



Approximation de l'Information Mutuelle basée sur le développement d'Edgeworth: application au recalage d'images médicales.

Mathieu Rubeaux

► To cite this version:

Mathieu Rubeaux. Approximation de l'Information Mutuelle basée sur le développement d'Edgeworth: application au recalage d'images médicales.. Traitement du signal et de l'image [eess.SP]. Université Rennes 1, 2011. Français. NNT: . tel-00632128

HAL Id: tel-00632128

<https://theses.hal.science/tel-00632128>

Submitted on 13 Oct 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



THÈSE / UNIVERSITÉ DE RENNES 1
sous le sceau de l'Université Européenne de Bretagne

pour le grade de
DOCTEUR DE L'UNIVERSITÉ DE RENNES 1

Mention : Traitement du Signal et Télécommunications

École doctorale MATISSE

présentée par

Mathieu RUBEAUX

préparée à l'unité de recherche LTSI - INSERM UMR 642

Laboratoire de Traitement du Signal et de l'Image
UFR Informatique Electronique (ISTIC)

**Approximation de
l'Information Mutuelle
basée sur
le développement
d'Edgeworth :
application au recalage
d'images médicales.**

**Thèse soutenue à Rennes
le 11 Juillet 2011**

devant le jury composé de :

Su RUAN

Professeur des universités / *rapporteur*

Dimitris VISVIKIS

Docteur INSERM / *rapporteur*

Jean-Pierre COCQUEREZ

Professeur des universités / *examinateur*

Mireille GARREAU

Professeur des universités / *co-directeur*

Jean-Claude NUNES

Maître de conférences / *co-directeur*

Laurent ALBERA

Maître de conférences / *co-directeur*

Remerciements

Je voudrais tout d'abord sincèrement remercier l'ensemble du jury qui a bien voulu accepter de prêter attention à mon travail de thèse. Leur regard neuf et leurs remarques pertinentes me sont d'ores et déjà très utiles. Je tiens tout particulièrement à remercier Mme Su Ruan et M. Dimitris Visvikis qui ont accepté la lourde tâche de rapporteur et qui ont dû relire mon travail dans des délais très court ! Merci également à Jean-Pierre Cocquerez, président du jury de thèse, qui par son expérience a apporté un point de vue intéressant à mes travaux.

Je tiens également à remercier MM. Jean-Louis Coatrieux et Lotfi Senhadji, directeurs du LTSI, qui oeuvrent au quotidien pour que nous puissions travailler dans les meilleures conditions possibles.

Un grand merci aussi à Mireille Garreau, Jean-Claude Nunes et Laurent Albera pour m'avoir encadré, ce qui ne doit pas être simple je l'avoue, pendant ces années. Merci à Laurent, qui avec Amar, a apporté les idées issues du traitement du signal qui sont à la base de ce travail de thèse. Merci à Mireille et Jean-Claude pour leur disponibilité à mon égard et leur engagement total quand cela était nécessaire. Je n'oublierai notamment pas tous les efforts que vous avez fournis au cours des dernières semaines, quand le temps commençait à jouer contre nous, et que je vous envoyais mes chapitres en flux tendu. Sans votre attention particulière et votre réactivité, cette fin de thèse aurait pu s'éterniser.

Ces années de thèse n'auraient également pas été les mêmes sans tous les collègues du LTSI que j'ai eu la chance de côtoyer. Merci donc à tous les membres du laboratoire, ceux que j'ai juste croisé à la cafétéria, ceux qui m'ont apporté leur aide pour faire avancer les travaux, ceux avec qui j'ai eu le plaisir d'enseigner. Merci en particulier à mes différents voisins de bureau qui ont agrémenté le quotidien de débats scientifiques (ou non..), et notamment à Amar et Benham. Merci aussi aux sportifs et autres coincheurs qui ont su rendre ces années très sympathiques pendant mais également en dehors des heures de travail : Monsieur T, MP, Carole, Gwen, les 2 Guillaumes, Simon, Alex, Mumu, Julien, Paul et Jérôme.

Je n'oublie pas non plus mes amis rennais, et notamment les poloïstes et compères du Saint-Georges qui sont toujours là quand une bonne occasion se présente.

Enfin et pour terminer, je tiens à remercier mes parents et mon frère qui m'ont supporté, au propre comme au figuré, pendant toutes ces années. Je fais également un petit coucou aux neveux Marin et Louis qui par leurs questions parfois impromptues poussent plus loin mes réflexions scientifico-philosophiques. Merci aussi à Julie qui m'a beaucoup épaulé dans la dernière ligne droite, à moi désormais de lui rendre la pareille.

Table des matières

Remerciements	1
Table des matières	2
Acronymes	7
Notations	9
Introduction	11
1 Revue bibliographique des techniques de recalage d’images médicales . .	13
1.1 Motivations cliniques	13
1.1.1 Recalage monomodal monopatient	14
1.1.2 Recalage multimodal monopatient	14
1.1.3 Recalage mono- ou multimodal multipatient	15
1.1.4 Recalage au LTSI et positionnement du problème	15
1.2 Recalage : un problème d’estimation paramétrique de mouvement	16
1.3 Classification des méthodes de recalage	17
1.3.1 Approches géométriques	19
1.3.2 Approches iconiques	20
1.3.3 Approche mixte	21
1.4 Mesures de similarité en recalage iconique	22
1.4.1 Modèle probabiliste des images	22
1.4.2 Conservation de l’intensité	23
1.4.3 Dépendance linéaire ou affine	24
1.4.4 Dépendance fonctionnelle	24
1.4.5 Dépendance statistique	26
1.5 Modèles de déformation	26
1.5.1 Transformations rigides	29
1.5.2 Transformations affines	29
1.5.3 Transformations non-rigides	30
1.5.3.1 Approches non-paramétriques	30
1.5.3.2 Approches paramétriques	31
1.6 Stratégie d’optimisation	35
1.6.1 Choix du pas	37
1.6.2 Choix de la direction de recherche	39
1.6.2.1 Méthode ne nécessitant pas de calcul de dérivées	40
1.6.2.2 Méthodes utilisant les dérivées premières	40
1.6.2.3 Méthode utilisant la dérivée seconde	42

1.7	Multirésolution et approches hiérarchiques	43
1.7.1	Complexité du modèle de déformation	43
1.7.2	Complexité des données	45
1.7.3	Approche bi-pyramidale	45
1.8	Conclusion	45
	Bibliographie	47
2	Information Mutuelle en recalage d'images médicales	57
2.1	Entropie de Shannon	57
2.2	Histogramme conjoint et entropie conjointe	58
2.3	Information mutuelle	60
2.4	Méthodes d'estimation de référence	61
2.4.1	Estimateur à histogramme	61
2.4.2	Estimateur à noyau	62
2.4.3	Estimateur à noyau gaussien	63
2.4.4	Estimateur à noyau B-Spline	65
2.5	Limitations de l'IM et solutions existantes	67
2.5.1	IM sensible au recouvrement partiel	67
2.5.2	IM incorporant une information spatiale	69
2.5.3	IM se prémunissant du calcul de l'histogramme conjoint	72
2.5.4	IM basées sur une autre mesure de l'entropie	77
2.6	Discussion au sujet des différentes mesures	79
2.7	Conclusion	80
	Bibliographie	81
3	Approximations de l'Information Mutuelle basées sur les cumulants	87
3.1	Prérequis mathématiques	87
3.1.1	Polynômes et tenseurs d'Hermite	88
3.1.1.1	Polynômes d'Hermite	88
3.1.1.2	Tenseurs d'Hermite	89
3.1.2	Outils statistiques : Moments et Cumulants	90
3.1.2.1	Moments et cumulants d'une variable aléatoire réelle	90
3.1.2.2	Extension au cas multidimensionnel	92
3.1.2.3	Propriétés des moments et cumulants	93
3.1.2.4	Estimation des cumulants	94
3.2	Cumulants en traitement du signal	94
3.2.1	Première ébauche	94
3.2.2	Développements de Gram-Charlier et d'Edgeworth	95
3.2.3	Analyse en Composantes Indépendantes et Information Mutuelle	96
3.2.4	Remarques sur cette approximation de l'IM	100
3.3	Calcul de nouvelles mesures basées sur un développement d'Edgeworth	100
3.3.1	Nouvelles hypothèses de travail	101
3.4	Conclusion	103
	Bibliographie	103
4	Résultats d'évaluation des mesures	105
4.1	Présentation des données	105
4.2	Critères d'évaluation	106
4.3	Développements logiciels	107

4.4	Evaluation des différentes mesures de similarité	109
4.4.1	Régularité pour une transformation rigide	111
4.4.2	Régularité face au bruit	118
4.5	Evaluations des optimiseurs	123
4.5.1	Présentation des paramètres des différents optimiseurs	123
4.5.2	Evaluation	126
4.5.3	Interprétation des résultats	127
4.6	Comportement de l'optimiseur en fonction du nombre d'échantillons	131
4.6.1	Interprétation	132
4.7	Evaluation du recalage rigide	135
4.7.1	Schéma monorésolution	136
4.7.2	Schéma multirésolution	137
4.8	Evaluation du recalage non-rigide	139
4.8.1	Base de données du laboratoire	139
4.8.2	Base de données EMPIRE	145
4.9	Conclusion	147
	Bibliographie	148
Conclusion		165
Annexes		171
A Transformations rigides et affines d'une image		171
B B-Splines		173
	Bibliographie	175
C Développements mathématiques		177
C.1	Approximation de l'entropie marginale par un développement d'Edgeworth . . .	177
C.2	Approximation de l'entropie conjointe par un développement d'Edgeworth . . .	181
C.2.1	Démonstration à l'ordre 3	181
C.2.2	Extension de l'approximation à l'ordre 4	187
	Bibliographie	189
D Formulaire : polynômes et tenseurs d'Hermite		191
	Bibliographie	195
E Base de données RIRE		197
	Bibliographie	200
F Exemple de fichier de paramètres utilisé avec elastiX		201
Table des figures		203

Acronymes

Chapitre 1

IRM	Imagerie par Résonnance Magnétique
SPECT	Tomographie d'émission de photons simples
TEP	Tomographie d'Emission à Positons
CT	Scanner
MCB	Mise en Correspondance de Blocs
MAP	Maximum A Posteriori
TPS	Thin Plate Splines
SDC	Somme des Différences au Carré
EQM	Erreur Quadratique Moyenne
SDA	Somme des Différences en valeur Absolue
ICN	Coefficient d'inter-corrélation normalisé
CC	Coefficient de Corrélacion
CA	Critère d'Alpert
RC	Rapport de Corrélacion
IM	Information Mutuelle
IMN	Information Mutuelle Normalisée
DDL	Degrés De Liberté
FFD	Free Form Deformations
PDC	Points De Contrôle
RBF	Fonctions à base radiale
QN	Quasi-Newton
DFP	Davidon-Fletcher-Powell
BFGS	Broyden-Fletcher-Goldfarb-Shanno
LBFGS	Limited memory BFGS

Chapitre 2

KL	Divergence de Kullback-Leibler
VA	Variable Aléatoire
RCE	Rapport de Corrélacion Entropique
IMM	IM Modifiée

IMNM	IMN Modifiée
RCEM	Rapport de Corrélation Entropique Modifié
IMR	IM Régionale
IMC	IM Conditionnelle
IML	IM Locale
ESE	Entropie Survivante Exponentielle
ESEG	Entropie Survivante Exponentielle Généralisée
ECR	Entropie Cumulative Résiduelle
ECRC	Entropie Cumulative Résiduelle Croisée
ECRCS	Entropie Cumulative Résiduelle Croisée Symétrique
RIRE	Retrospective Image Registration Evaluation
RREP	Retrospective Registration Evaluation Project
IMMCS	IM Multi-Caractéristiques Symétrique

Chapitre 3

ACI	Analyse en Composantes Indépendantes
H_{3E}	Entropie conjointe basée sur un développement d'Edgeworth à l'ordre 3
H_{3E}	Entropie conjointe basée sur un développement d'Edgeworth à l'ordre 4
3EIM	IM basée sur un développement d'Edgeworth à l'ordre 3
4EIM	IM basée sur un développement d'Edgeworth à l'ordre 4
3EIMN	IMN basée sur un développement d'Edgeworth à l'ordre 3
4EIMN	IMN basée sur un développement d'Edgeworth à l'ordre 3

Chapitre 4

IMN_H	IMN estimée par Histogramme
IM_G	IM estimée par fenêtrage de Parzen avec noyau Gaussien
IN_{BS}	IM estimée par fenêtrage de Parzen avec noyau B-Spline
IMN_{BS}	IMN estimée par fenêtrage de Parzen avec noyau B-Spline
IMOP	utilité de l'Imagerie Médicale dans l'Optimisation de la Pose de prothèses cardiaques implantables
MICCAI	Medical Image Computing and Computer Assisted Intervention
EMPIRE	Evaluation of Methods for Pulmonary Image Registration
ITK	Insight ToolKit
G	descente de Gradient
GC	Gradient Conjugué non linéaire
GS	Gradient Stochastique
FR	Facteur de Relaxation

Notations

\boxtimes	Opérateur du produit de Hadamard
$^{\top}$	Transposée
$E[.]$	Espérance mathématique
$Cov(.)$	Covariance
$Var(.)$	Variance
$tr(.)$	Trace d'une matrice
$H(\boldsymbol{\mu}_k)$	Matrice hessienne
$D^{(1)}$	Ensemble des niveaux de gris de l'image de référence
$D^{(2)}$	Ensemble des niveaux de gris de l'image de référence
$\boldsymbol{\nu} = (\nu_x, \nu_y, \nu_z)^{\top}$	Pixel de l'image
\mathcal{V}	Ensemble des pixels pris en compte dans l'estimation
$x^{(1)}$	Variable aléatoire associée à l'image de référence
$x^{(2)}$	Variable aléatoire associée à l'image flottante
\boldsymbol{x}	Vecteur aléatoire des 2 variables aléatoires $x^{(1)}$ et $x^{(2)}$
$x^{(1)}(\boldsymbol{\nu})$	Image de référence
$x^{(2)}(\boldsymbol{\nu})$	Image flottante
$p_{x^{(1)}}$	Densité de probabilité marginale de $x^{(1)}$
$p_{x^{(2)}}$	Densité de probabilité marginale de $x^{(2)}$
$p_{x^{(1)} x^{(2)}}$	Densité de probabilité conditionnelle
$m_{x^{(1)} x^{(2)}}$	Espérance conditionnelle
$\sigma_{x^{(1)} x^{(2)}}^2$	Variance conditionnelle
$p_{\boldsymbol{x}}$	Densité de probabilité conjointe
$h_{\boldsymbol{x}}$	Histogramme conjoint de \boldsymbol{x}
$H(x^{(1)})$	Entropie marginale associée à l'image de référence
$H(x^{(2)})$	Entropie marginale associée à l'image flottante
$H(x^{(2)} x^{(1)})$	Entropie conditionnelle
$H(\boldsymbol{x})$	Entropie conjointe de \boldsymbol{x}
$\phi_x(u)$	Distribution normale univariée
$\phi_{\boldsymbol{x}}(\boldsymbol{u})$	Distribution normale bivariée

$\kappa_{\mathbf{x}}$	Matrice de covariance de \mathbf{x}
$\kappa_{\mathbf{x}}^{i,j}$	Cumulant d'ordre 2 de \mathbf{x}
$\kappa_{\mathbf{x}}^{-1}$	Inverse de $\kappa_{\mathbf{x}}$
$\rho = \kappa_{\mathbf{x}}^{1,2}$	Covariance entre $x^{(1)}$ and $x^{(2)}$
$H_k(u)$	k -ème polynôme d'Hermite
$H_{i_1 \dots i_k}$	Tenseur d'Hermite covariant
μ_q	Moment d'ordre q de x
κ_q	Cumulant d'ordre q de x
\mathcal{P}_k	Point de contrôle d'indice k
K	Nombre total de points de contrôle
$\xi^{(i)}$	Vecteur de déplacement d'un point de contrôle
$\beta^{(n)}(z)$	B-Spline de degré n
$\beta^{(n)}(\boldsymbol{\nu})$	Vecteur de B-Splines de degré n
$\boldsymbol{\mu}$	Vecteur de paramètres à estimer
$\boldsymbol{\mu}_0$	Vecteur de paramètres optimal
\mathbf{a}_k	Pas de l'optimiseur
\mathbf{d}_k	Direction de recherche
$\mathbf{g}(\boldsymbol{\mu}_k)$	Gradient de la fonction de coût évalué à la position $\boldsymbol{\mu}_k$
$\tilde{\mathbf{g}}_k$	Approximation de $\mathbf{g}(\boldsymbol{\mu}_k)$

Introduction

L'imagerie médicale est actuellement au coeur du dispositif permettant au praticien d'établir un diagnostic ou de planifier une thérapie. Les avancées technologiques du XXème siècle ont en effet donné naissance à de nombreuses modalités d'imagerie, qui autorisent une vision toujours plus précise de l'anatomie des patients. En parallèle de ces progrès, la nécessité d'analyser les images issues de ces modalités s'est accrue, car les volumes de données à traiter sont très importants. Faute d'outils efficaces d'analyse automatique, l'exploitation des données est souvent faite manuellement.

Le recalage d'images tient une place très importante dans l'analyse et l'interprétation de ces données. Il permet, dans le cadre du diagnostic, de mettre en correspondance des images provenant de plusieurs modalités d'imagerie, de suivre le traitement d'un patient au cours du temps ou de réaliser des analyses inter-patients. Il permet également dans le cadre de la conduite de gestes interventionnels guidés par l'image, d'intégrer dans un même espace des informations issues d'images acquises en situations pré- et per-opératoires. L'utilisation des techniques de recalage a été rendue accessible par les progrès récents de l'informatique, et ces techniques autorisent aujourd'hui la combinaison d'informations provenant de différents imageurs. Néanmoins, le besoin de méthodes de recalage efficaces est toujours important à l'heure actuelle.

Dans ce travail, nous nous sommes concentrés sur une des composantes des méthodes de recalage qui est la mesure de similarité utilisée pour mettre en correspondance les images. Nous nous sommes plus particulièrement intéressés à l'approche iconique, qui prend en compte uniquement l'information portée par les intensités des pixels des images à recaler. Cette approche présente l'avantage d'être totalement automatique, puisqu'aucune segmentation préalable n'est nécessaire.

L'apport principal de ce travail de thèse est de proposer de nouvelles mesures de similarité basées sur les cumulants et le développement d'Edgeworth, et qui pour certaines d'entre elles approximent l'Information Mutuelle, qui est une mesure de similarité de référence en recalage. Les tests menés sur ces nouvelles mesures montrent leur efficacité pour le recalage d'images médicales. De plus, la généralité de l'approche proposée autorise l'utilisation de ces mesures dans des situations variées. Nous détaillons maintenant l'organisation de ce manuscrit.

Le premier chapitre présente une revue bibliographique des techniques utilisées pour le recalage d'images médicales. L'accent est volontairement mis sur les approches iconiques, qui considèrent l'information apportée par l'intensité des pixels des images à recaler. Au cours de cette étude, nous guidons progressivement le lecteur vers la finalité de ce travail, en détaillant particulièrement les différentes composantes des méthodes de recalage utilisées au cours de cette thèse. Nous montrons également le cheminement qui a conduit les chercheurs à considérer l'Information Mutuelle (IM) comme la mesure de référence pour le recalage d'images médicales multimodales. Cette mesure de similarité étant au coeur de ce travail, nous y consacrons un

chapitre.

En effet, le deuxième chapitre s'intéresse particulièrement à l'IM. Cette mesure de dépendance, qui a vu le jour au milieu du XXème siècle, est utilisée pour le recalage d'images médicales depuis une quinzaine d'années. Dans un premier temps, nous expliquons les techniques classiques d'estimation de l'IM qui ont été développées. Puis, nous mettons en avant les qualités et défauts intrinsèques de ces estimateurs. Nous présentons ensuite les améliorations qui ont été apportées à ces estimateurs, ainsi que les nouvelles méthodes employées pour construire une approximation de l'IM. Ce chapitre justifie ainsi la nécessité de se doter de nouveaux estimateurs de l'IM dans le cadre du recalage.

Le chapitre trois présente les mesures de similarité que nous avons développées au cours de cette étude. Nous commençons par donner les outils mathématiques nécessaires à la compréhension du chapitre. Parmi ceux-ci, on compte les cumulants, des statistiques qui connaissent un important succès en traitement du signal. Nous avons choisi d'utiliser ces cumulants pour construire des mesures de similarité exploitables dans le cadre du recalage. Pour ce faire, nous nous sommes servis du développement d'Edgeworth, qui permet d'approximer une densité de probabilité à partir des cumulants des variables aléatoires considérées. A l'aide de ce développement d'Edgeworth, nous avons conçu des approximations des entropies marginales et conjointes de deux variables aléatoires potentiellement corrélées, ce qui nous permet de proposer des estimateurs de l'IM basés sur les cumulants.

Le quatrième et dernier chapitre présente les tests conduits pour valider l'utilisation de nos mesures de similarité dans le cadre du recalage rigide et non-rigide d'images médicales multimodales. Ces tests ont été menés de façon progressive. Au départ, nous nous sommes intéressés aux qualités propres de ces mesures, en comparaison avec les mesures de référence. Puis nous avons intégré la mesure la plus efficace dans un schéma global de recalage. Une attention particulière a été portée sur le paramétrage des différentes composantes de cette méthode globale. Finalement, disposant d'une méthode de recalage utilisant notre mesure de similarité, nous avons testé son efficacité sur plusieurs bases de données. Nous avons considéré des tâches de recalage rigide et non-rigide, pour des images mono- et multimodales, et dans des contextes applicatifs différents. La diversité des situations rencontrées permet de considérer que la mesure proposée est assez générique pour être utilisée dans un grand nombre de problèmes nécessitant un recalage.

Les annexes viennent enfin appuyer notre propos. Les plus importantes sont les annexes C et D qui présentent en détail les développements mathématiques que nous avons conduits pour proposer ces nouvelles mesures de similarité.

Les perspectives sont nombreuses pour ce travail de thèse, nous les expliquons dans les différents chapitres et les synthétisons dans la conclusion de ce manuscrit.

Chapitre 1

Revue bibliographique des techniques de recalage d'images médicales

Depuis trois décennies, l'imagerie médicale a connu un essor très important lié aux avancées technologiques dans le cadre de l'aide au diagnostic ou de l'aide à la thérapie. Il existe aujourd'hui une multitude de modalités d'imagerie médicale (radiologie conventionnelle, tomodensitométrie, Imagerie par Résonnance Magnétique (IRM), échographie, scintigraphie, tomographie d'émission de photons simples (SPECT), Tomographie d'Emission à Positons (TEP), etc.), chacune apportant une information spécifique et complémentaire. Ces images, qui peuvent être de nature bi- ou tri-dimensionnelles et auxquelles la composante temporelle peut venir s'ajouter, représentent une masse de données très importante à analyser. Pourtant, ces données sont encore souvent exploitées manuellement, faute d'outils efficaces d'analyse semi-automatique ou automatique.

Le recalage d'images joue un rôle central dans l'exploitation et l'interprétation de ces images. En effet, il permet de comparer, voire de fusionner les données issues de différentes modalités d'imagerie ou de patients différents pour un examen donné, mais également le suivi de patient ou l'analyse inter-patients. Un intérêt croissant a été porté à ce domaine ces dernières années. On peut y voir deux explications : tout d'abord, les modalités d'acquisition d'imagerie médicale s'étant diversifiées, il est intéressant de pouvoir combiner ces images pour extraire d'avantage d'informations. Ensuite, les progrès de l'informatique rendent désormais possible des calculs inenvisageables il y a encore quelques années.

L'objectif de cet état de l'art n'est pas de recenser de manière exhaustive toutes les méthodes de recalage aujourd'hui disponibles. Nous nous emploierons plutôt à définir la problématique essentielle du recalage d'images, problématique proche de l'estimation de mouvement. Par souci de concision, nous réduirons progressivement le champ des méthodes disponibles pour se focaliser sur celles correspondant à notre sujet d'étude, à savoir les méthodes de recalage iconiques (c'est à dire sans segmentation préalable) basées sur l'information mutuelle.

1.1 Motivations cliniques

En quelques mots, le recalage d'images consiste à déformer par une transformation géométrique (ou spatiale) une image dite flottante pour la faire correspondre parfaitement à une autre, dite de référence. Sous ce principe simple se côtoient de nombreuses méthodes, avec des finalités cliniques différentes. Pour mieux appréhender les enjeux du recalage, nous allons commencer par donner un aperçu des applications médicales envisageables.

1.1.1 Recalage monomodal monopatient

C'est le cas le plus simple. Ici, les images proviennent du même patient, prises avec la même modalité d'imagerie, mais à des instants différents. Les usages que l'on peut faire de ces images comprennent notamment :

- **le contrôle d'opérations chirurgicales** : une comparaison des images pré- et postopératoires permet de vérifier que l'acte chirurgical a atteint le résultat escompté (ablation complète d'une tumeur, par exemple).
- **le suivi de patient** : des images ont été acquises sur différentes périodes, et sont comparées entre-elles pour juger de l'évolution de la pathologie ou pour rendre compte de l'efficacité d'un traitement.
- **l'estimation de mouvement cardiaque** : l'imagerie cardiaque permet d'évaluer la fonction cardiaque, d'un point de vue morphologique, dynamique et/ou fonctionnel. Typiquement, des images tomodensitométriques bi- ou tridimensionnelles (scanner, IRM, échographiques) sont acquises à différents instants du cycle cardiaque. Le but du recalage est, ici, d'estimer le mouvement du coeur à partir de ces séquences d'images.
- **la compensation du mouvement respiratoire** : lors de l'acquisition d'images médicales, le mouvement respiratoire du patient peut induire des artefacts sur les images ou des différences importantes entre des images prises à des instants différents. Le recalage peut ainsi être utilisé pour compenser ce mouvement respiratoire.

1.1.2 Recalage multimodal monopatient

Cette démarche est plus complexe puisqu'il s'agit de mettre en correspondance des images qui peuvent être de natures différentes. Les applications typiques sont :

- **la fusion de données** : elle résulte de la combinaison de plusieurs modalités d'imagerie. Chaque modalité d'imagerie possède ses caractéristiques propres, et apporte une information différente sur la pathologie ou l'organe considéré. Les modalités d'imagerie peuvent se décomposer en deux grandes familles complémentaires et spécifiques :
 - imagerie anatomique (radiologie conventionnelle, scanner ou tomodensitométrie, IRM, échographie, etc).
 - imagerie fonctionnelle (scintigraphie, PET, SPECT, échographie Doppler, IRM fonctionnelle, IRM de diffusion, etc).

Aussi, il peut parfois être intéressant de fusionner ces différentes informations pour établir un meilleur diagnostic. La fusion de données va donc consister à recaler des images de modalités différentes. C'est un domaine en plein essor, notamment grâce aux avancées technologiques qui permettent de traiter des masses de données toujours plus importantes. La finalité est évidemment d'apporter un diagnostic plus précis, puisque se basant sur plusieurs types d'imageries complémentaires.

- **la planification d'intervention** : grâce à une meilleure connaissance de l'anatomie interne d'un patient, les médecins peuvent établir un meilleur diagnostic et ainsi mieux planifier leur intervention. En effet, les organes par exemple peuvent être de tailles et de formes différentes suivant les patients. Par une analyse d'images pré-opératoires, les médecins ont ainsi accès à une description anatomique précise de la cible, des organes à risques avoisinant la cible ou des risques liés à l'insertion d'instruments chirurgicaux. Il est important que chacune de ces descriptions soit précise, et la combinaison d'informations issues de modalités d'imagerie différentes permet d'améliorer la décision finale.

1.1.3 Recalage mono- ou multimodal multipatient

De nombreuses études évoquent la possibilité de comparer des données issues de plusieurs patients. Cela permet, par exemple, de détecter des pathologies par comparaison avec un groupe d'individus sains. Généralement, la solution retenue consiste à déformer des images individuelles pour les faire correspondre à un atlas standard. Par exemple, en imagerie cérébrale fonctionnelle, la création d'un atlas numérique consiste à moyenner les images d'une même modalité provenant de plusieurs patients sains (ceci implique une première étape de recalage inter-patients). Une carte, ou modèle, de l'anatomie fonctionnelle du cerveau humain est ainsi obtenue. En recalant l'image du cerveau d'un patient sur cet atlas, il est possible de repérer la zone du cerveau lésée. L'utilisation des atlas numériques concernent de nombreuses applications, parmi lesquelles nous pouvons notamment citer :

- l'aide au diagnostic, par comparaison d'un individu à une population « terrain » représentée par un atlas.
- la comparaison de deux groupes d'individus entre-eux.
- l'apport d'information haut niveau. En effet, les atlas peuvent être utilisés comme source d'information pour des opérations telles que la segmentation, le recalage, la reconstruction, le suivi de lésions au cours du temps, etc.

1.1.4 Recalage au LTSI et positionnement du problème

Ces quelques exemples permettent d'entrevoir la place très importante qu'occupe le recalage dans l'analyse des images médicales. Au Laboratoire Traitement du Signal et de l'Image, les problématiques de recalage rencontrées sont nombreuses. Nous donnons ici quelques exemples, afin de situer notre travail.

Tout d'abord, des travaux portent sur l'estimation du mouvement cardiaque à partir de séquences d'images tridimensionnelles. Cette estimation est un enjeu de premier ordre dans la détection des cardiopathies, et elle a bénéficié des progrès réalisés en imagerie dynamique, avec notamment le développement de l'échographie tridimensionnelle et du scanner multibarettes. Dans ce contexte, des méthodes de recalage dédiées aux volumes scanners 3D ont été proposées [Simon, 2005, Simon et al., 2005, Simon et al., 2006, Garreau et al., 2006].

Toujours dans le domaine de l'imagerie cardiaque, la thérapie de resynchronisation cardiaque, geste interventionnel difficile, peut être optimisé par une identification précise des sites de stimulation du ventricule gauche (VG), à partir d'informations anatomiques et électriques issues respectivement d'imagerie scanner multibarette et Cartographie Electro-Anatomique (CEA). Afin de représenter ces données dans un même environnement, une méthode de fusion d'informations en deux temps est proposée : (1) recalage rigide de surfaces ; (2) fusion de surface 3D et de cartographie électrique. Dans la première étape, deux surfaces 3D du VG obtenues à partir de la CEA et d'images scanner segmentées sont recalées de façon automatique ou semi-automatique. Dans la deuxième étape, les délais d'activation électrique (DAE) en chaque point de l'endocarde sont estimés depuis la CEA. Enfin, une interface graphique permet de visualiser les DAE sur le VG. Des travaux récents [Tavard et al., 2009, Tavard et al., 2010] proposent des méthodes de recalage/fusion d'information dans ce contexte applicatif.

Dans le cadre du traitement du cancer de la prostate par radiothérapie guidée par l'image, le recalage non-rigide tient également une place importante au laboratoire. Les motivations pour l'utilisation du recalage sont multiples. Tout d'abord, la prédiction de toxicité est un enjeu fondamental, car elle permet de mieux contrôler les radiations reçues par le patient. Dans ce cadre, la mise au point d'un modèle statistique de toxicité permettrait de mettre en évidence des relations entre la dose distribuée et la survenue de toxicité. La construction de ce

modèle passe par la mise au point d'un atlas, construit par recalage inter-patient, permettant le mapping de dose sur un modèle issu de cet atlas [Dréan et al., 2011, Acosta et al., 2011]. D'autre part, lorsqu'on essaie de faire un calcul de la dose cumulée reçue en cours de traitement, il est intéressant de pouvoir suivre les mouvements et la déformation des organes tout au long de ce traitement, pour savoir quelle dose de radiation est effectivement reçue par le patient. Cela apporte une aide précieuse à la planification des séances de radiothérapie. Dans ce contexte, des méthodes de recalage ont déjà été appliquées avec succès [Cazoulat et al., 2011].

Quelle que soit la finalité clinique, les méthodes de recalage d'images médicales font apparaître des enjeux communs. Souvent, les contraintes imposées par les tissus ou organes observés nécessitent la prise en compte de déformations complexes et non-rigides. C'est notamment le cas dans les exemples précédemment cités. De plus les volumes de données à analyser, de par les progrès de imageurs, sont toujours plus importants. Il faut donc se doter de méthodes de recalage efficaces pour traiter toutes ces images en un minimum de temps, la finalité étant un diagnostic rapide.

Nous le voyons, les besoins en méthodes de recalage sont importants au laboratoire, et plus généralement dans la communauté médicale. Dans ce travail de thèse, nous ne nous sommes pas concentrés sur une problématique médicale particulière, mais nous avons plutôt cherché à proposer de nouveaux outils permettant un recalage non-rigide et/ou multimodal de qualité. En particulier, nous nous sommes intéressés aux mesures de similarité utilisées en recalage d'images médicales multimodales et proposons de nouvelles méthodes pour approximer ces mesures. De par leur généralité, ces méthodes sont susceptibles d'être utilisées dans de nombreux domaines d'application nécessitant un recalage. Nous verrons au chapitre 4 que nous les avons déjà testées sur plusieurs problèmes cliniques particuliers, parmi lesquels la construction d'atlas inter-patients auquel nous faisons référence précédemment. Intéressons nous maintenant aux différentes méthodologies de recalage qui ont été développées dans la littérature.

1.2 Recalage : un problème d'estimation paramétrique de mouvement

Le recalage peut, avant tout, être vu comme un problème d'estimation de mouvement dans une séquence d'images, un thème de recherche fondamental en traitement d'images. Les domaines d'application de l'estimation de mouvement vont de la compression vidéo à l'interprétation des images satellitaires en passant, et c'est notre sujet dans ce travail, par la caractérisation d'organes dynamiques. Dans la littérature, on distingue deux grands types de techniques : les techniques non-paramétriques et les techniques paramétriques. Il existe trois grands types de méthodes non-paramétriques d'estimation de mouvement :

- **les méthodes différentielles**. Elles visent à estimer le flux optique entre deux images à partir de l'estimation de gradients spatio-temporels de l'intensité en chacun des pixels de l'image. Elles sont très utilisées en traitement d'images depuis l'article originel de Horn et Schunk [Horn and Schunk, 1981]. Ces méthodes sont en général utilisées pour mesurer de petits déplacements, et font l'hypothèse d'une conservation de l'intensité entre les images.
- **les méthodes de Mise en Correspondance de Blocs (MCB)**. Leur principe est de découper une image de référence en blocs de pixels ou de voxels. Le but est ensuite d'essayer de trouver dans l'image suivante, pour chacun des blocs initiaux, le bloc le plus ressemblant. L'hypothèse admise est que le mouvement entre deux images est localement translationnel. Ces méthodes sont très utilisées, notamment pour la compression d'images ou de vidéos (MPEG) [Stiller and Konrad, 1999].

- **les méthodes statistiques.** Parmi les méthodes statistiques, les méthodes Markoviennes ou Bayésiennes sont les plus usitées. Elles reposent sur une formulation probabiliste du champ de mouvement. Le champ considéré est vu comme étant une réalisation particulière d'un champ aléatoire. La formulation la plus répandue repose sur l'estimateur du Maximum A Posteriori (MAP) [Gee et al., 1995, Odobez and Bouthemy, 1995, Stiller and Konrad, 1999].

Nous ne présenterons pas en détails les méthodes non-paramétriques, puisque ce n'est pas le sujet de cette thèse. En effet, le recalage suppose connu un certain modèle paramétrique du mouvement ou du déplacement, puisqu'il se fonde sur l'estimation des paramètres d'une transformation modélisant le champ de mouvement. Le reste de ce document s'intéresse donc à l'estimation de mouvement paramétrique.

1.3 Classification des méthodes de recalage

Au vu du nombre d'approches recensées dans le littérature, il apparaît utopique d'être exhaustif. Souvent, les états de l'art se veulent spécifiques à un problème donné, et nous suivrons donc cette voie. En 1992, Lisa Gottesfield Brown propose un état de l'art devenu depuis une référence [Brown, 1992] dans lequel elle propose la première classification des méthodes de recalage, dans un cadre général. D'autres auteurs ont depuis apporté des éclairages différents et complémentaires : l'article de Zitova [Zitova and Flusser, 2003] est dans la continuité de [Brown, 1992] et apporte en quelque sorte une mise à jour nécessaire, avec un recul de 10 ans sur le sujet. Le cadre est ici toujours le recalage en général. D'autres états de l'art s'attachent spécifiquement au recalage d'images médicales. C'est le cas notamment de [Maintz and Viergever, 1998] qui recense plus de 300 références et qui définit une nouvelle classification des méthodes de recalage basée sur 9 critères (dimension des images, nature du recalage, nature de la transformation, domaine de la transformation, etc.), mais aussi de [Gilardi, 1996, Lester and Arridge, 1999, Barillot, 1999, Hill et al., 2001, Makela et al., 2002, Hutton and Braun, 2003, Pluim et al., 2003, Crum et al., 2004]. Le lecteur intéressé au développement logiciel des techniques de recalage pourra se référer à [Shams et al., 2010], qui détaille notamment l'utilisation des processeurs graphiques et multi-coeurs dans ce cadre.

De nombreuses thèses viennent compléter ce panorama : [Mangin, 1995, Roche, 2001] [Kybic, 2001, Sarrut, 2000, Cachier, 2002, Atif, 2004, Noblet, 2006] concernent le recalage d'images cérébrales, problème pour lequel un modèle de déformation affine est souvent retenu. Nous verrons dans ce chapitre que ce modèle est relativement simple par rapport aux modèles non-rigides nécessaires pour recalcr des organes mous tels les poumons, le coeur etc. La thèse [Loeckx, 2006] est également très intéressante puisqu'elle traite du recalage par les B-Splines, notion qui sera développée dans la suite. Pour terminer, nous invitons le lecteur à lire les thèses de [Petitjean, 2003, Simon, 2005, Loeckx, 2006, Delhay, 2006, Chenoune, 2008, Mora Cofre, 2008] qui étudient le mouvement du coeur. La première de celles-ci [Petitjean, 2003] utilise une approche iconique comparable à la notre, tandis que les autres études proposent un recalage géométrique. Nous allons voir en détail par la suite les différences que cela implique.

Pour cet état de l'art, nous avons choisi la classification de Brown [Brown, 1992] qui, bien que ne comportant que 4 catégories, va nous permettre de mettre en avant les enjeux de cette thèse. En effet, si nous formalisons le processus de recalage, il peut-être vu comme un problème de minimisation d'une fonctionnelle :

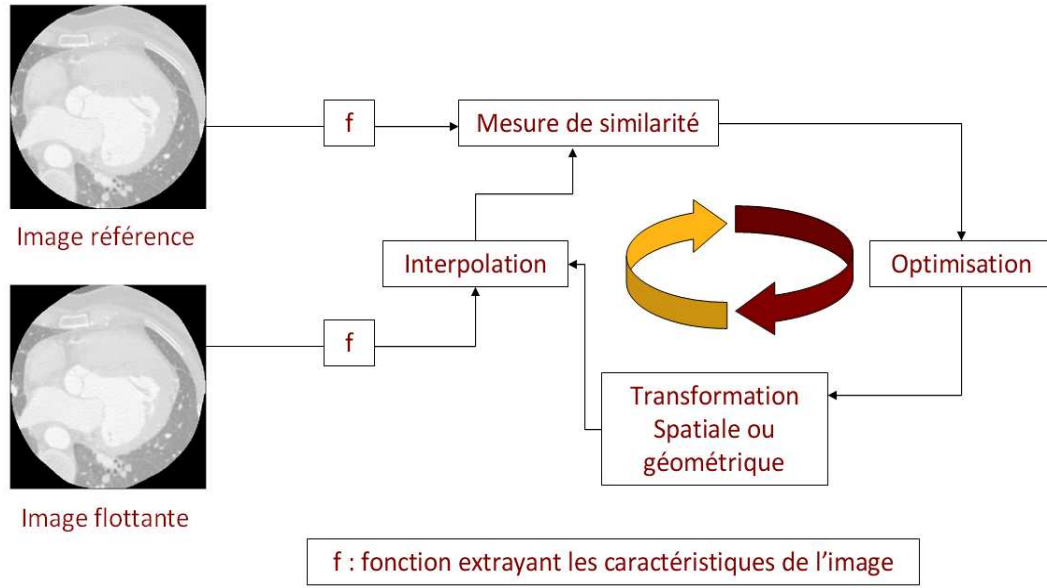


FIGURE 1.1 – Principe du recalage.

$$\tilde{T} = \arg \min_T S(f(x^{(1)}(\nu)), f((x^{(2)}(T(\nu)))) \quad (1.1)$$

où les critères retenus apparaissent :

- $x^{(1)}(\nu)$ et $x^{(2)}(\nu)$ sont respectivement l'image de référence et l'image flottante.
- f est la fonction extrayant les caractéristiques de l'image. Deux grandes classes de méthodes se distinguent. Nous trouvons d'un côté celles qui utilisent l'ensemble de l'information des images (méthodes iconiques), de l'autre celles utilisant le résultat d'une étape préalable de segmentation (méthodes géométriques). Nous situerons notre travail du côté des approches iconiques.
- S représente la mesure de similarité. Celle-ci dépend des attributs de l'image utilisés, et nous expliquerons pourquoi nous nous sommes orientés vers l'information mutuelle. Nous nous appuyerons notamment sur [Pluim et al., 2003] qui propose un état de l'art du recalage d'images médicales basé sur celle-ci.
- T symbolise un élément de l'espace des transformations possibles, ou modèle de déformation. Les premières approches estiment des transformations rigides ou affines. Cependant dans le cadre de cette thèse, nous nous intéressons au recalage d'organes élastiques pour lequel un modèle non-rigide est indispensable. Un tel modèle offre plus de degrés de liberté (DDL) vis-à-vis de la transformation souhaitée, permettant l'estimation d'une transformation plus complexe mais plus proche de la réalité. L'article [Rohr, 2000] appuiera notre démonstration.
- $\arg \min_T$ induit la stratégie d'optimisation. Le volume de données à traiter étant généralement très important, il apparaît nécessaire de choisir judicieusement l'algorithme d'optimisation. Dans le cadre de notre travail, une étude menée dans [Klein et al., 2007] a conforté nos choix que nous expliciterons par la suite.

Cette présentation, bien que générale, s'orientera progressivement vers les décisions que nous avons prises au cours de ce travail de thèse. Tout d'abord, pour mener à bien le recalage de deux images entre-elles, il faut commencer par choisir l'information que l'on va extraire de ces

images. Trois grandes approches sont alors possibles :

- la première s’appuie sur des primitives géométriques extraites de l’image, comme des points, lignes ou surfaces caractéristiques. Ces méthodes sont donc communément appelées **méthodes géométriques**.
- la seconde exploite toute l’information brute contenue dans l’image (sans segmentation préalable), à savoir les niveaux de gris. On parle alors de **méthodes iconiques**.
- la troisième enfin combine les caractéristiques des deux approches précédentes. Ces méthodes sont regroupées sous l’appellation de **méthodes mixtes**.

Ces 3 approches ont chacune leurs intérêts et leurs limitations. Les méthodes géométriques seront présentées intentionnellement de façon succincte, l’objectif étant avant tout de donner un point de vue au lecteur et de pouvoir comparer les avantages et limites de chacune. En revanche, nous expliquerons plus précisément les méthodes iconiques qui rentrent dans le cadre de notre étude.

1.3.1 Approches géométriques

Dans ce type d’approche, des caractéristiques géométriques sont extraites des 2 images à recalcr, puis le recalage à proprement parler consistera à retrouver la transformation entre ces caractéristiques issues des 2 images. Cette transformation géométrique est celle qui minimise la distance entre les caractéristiques appariées. C’est historiquement la première méthode à avoir été utilisée, et ceci s’explique peut-être par le fait que le processus de recalage géométrique ressemble à ce qui est fait naturellement par la perception visuelle humaine : on cherche dans les images étudiées des points, lignes ou surfaces remarquables et communes que l’on essaie de mettre en correspondance. Les primitives utilisées peuvent être classées en fonction de leur nature [Mangin, 1995] : **primitives géométriques** (segments ou plans), **primitives topologiques** (composantes connexes, cavités, surfaces ou intersections de surfaces), **primitives morphologiques** (squelette ou enveloppe connexe), ou **primitives différentielles** (lignes de crête, discontinuités). Les attributs géométriques peuvent alors être des marqueurs externes (comme les points d’un cadre stéréotaxique visibles par les différents imageurs par exemple), des points, des contours ou des surfaces.

L’avantage majeur est un temps de traitement limité. En effet, en ne considérant qu’une représentation compacte de l’image, la charge de calcul s’en trouve d’autant réduite. De plus, les primitives extraites se révèlent très informatives puisqu’elles se basent sur des caractéristiques de haut niveau de l’image liées à l’objet d’intérêt.

Cependant, la majorité des approches se base soit sur une sélection manuelle, soit sur une détection semi-automatique des amers (ou primitives géométriques) communs aux deux images, ce qui impose un pré-traitement des données pour extraire ces points, lignes ou surfaces d’intérêt. Sur des structures rigides telles que les os, la tâche est relativement simple puisque les amers identifiables seront stables sur les différentes images. Le problème est par contre beaucoup plus complexe sur des structures molles telles que le coeur, puisque la sélection d’amers (surface endocardique, par exemple) nécessitera une segmentation préalable. De la qualité de cette segmentation initiale dépendra ensuite la robustesse du recalage. De plus, pour obtenir un résultat intéressant en recalage non-rigide, le nombre de caractéristiques géométriques devra être important, la qualité de l’interpolation des images entre ces caractéristiques dépendant directement de leur nombre.

Nous ne nous étendrons pas sur les distances utilisées pour apparier les primitives géométriques. Néanmoins nous pouvons donner quelques axes en fonction des primitives utilisées :

- pour des **points** (tels les points de Moravec ou les passages par zéro du Laplacien), la

Méthode	Géométrique	Iconique
Avantages	<ul style="list-style-type: none"> - Temps de traitement réduit. - Primitives très informatives. 	<ul style="list-style-type: none"> - Adapté au recalage multimodal. - Pas de prétraitement requis. - Méthode automatique.
Inconvénients	<ul style="list-style-type: none"> - Sélection manuelle ou semi-manuelle des primitives. - Généralement réservé au recalage monomodal. - Prétraitement nécessaire (segmentation). 	<ul style="list-style-type: none"> - Primitives peu informatives (de bas niveau). - Temps de traitement important.

TABLE 1.1 – Avantages et inconvénients des différentes approches en recalage d’images médicales

distance classiquement utilisée est une simple norme euclidienne. Elle conduit, dans le cadre du recalage rigide et affine, à une solution analytique par la méthode de Procrustes [Arun et al., 1987].

- pour des **courbes** (plus précisément l’orientation et la direction de ces courbes) ou des **surfaces** obtenues généralement par des modèles déformables ou level-set, plusieurs approches sont possibles : l’algorithme ICP (Iterative Closest Point) [Besl and McKay, 1992], Head and Hat [Pelizzari et al., 1989], les cartes de distances ou encore l’interpolation /approximation par fonctions radiales (RBF pour Radial Basis Function). Pour cette dernière approche, citons [Arad et al., 1994, Rohr et al., 1999] qui utilisent les splines de plaques minces (TPS pour Thin Plate Splines) comme fonction radiale, tandis que [Davis et al., 1997] se sert des splines élastiques ou encore [Nielson, 1993] des splines volumiques.

Le lecteur pourra se référer à l’état de l’art proposé dans [Velkamp and Hagedoorn, 1999].

Etant donnée l’hypothèse de travail initiale de cette thèse (recalage multimodal non-rigide), l’approche géométrique a été écartée dans le cadre de notre étude. Ce choix a été porté par plusieurs considérations. Tout d’abord, les méthodes géométriques nécessitent un prétraitement dont nous voulions nous prémunir afin d’obtenir une méthode totalement automatique. De plus, la détection de primitives dans des images de nature différentes demeure un problème délicat, et sa nécessité comme première étape d’une méthode de recalage peut s’avérer complexe. En effet, cette première étape peut entraîner des erreurs (dûes par exemple à une segmentation incorrecte) qui se répercutent directement sur la seconde étape qui est le recalage lui-même. Nous avons donc décidé d’utiliser une approche relevant des méthodes iconiques que nous présentons maintenant.

1.3.2 Approches iconiques

Les méthodes iconiques se fondent sur l’estimation d’une déformation qui maximise une fonction d’énergie traduisant la similitude entre les images. La mesure de similarité opère ici généralement sur la totalité des données brutes des images (niveaux de gris), ou plus rarement sur une représentation de ces niveaux de gris : ainsi, [Mellor and Brady, 2005] appliquent une transformée de Fourier sur les images pour travailler dans le domaine fréquentiel en utilisant l’information de phase, tandis qu’une décomposition de l’image en ondelettes est utilisée par [Xue and Davatzikos, 2004]. Les méthodes utilisant directement les niveaux de gris sont apparues dans les années 70 [Barnea and Silverman, 1972, McGillem and Svedlow, 1976,

[Svedlow et al., 1978] dans le cadre rigide et monomodal. Avec les progrès de l'imagerie médicale est ensuite apparu le problème du recalage multimodal qui se définit comme la comparaison d'images provenant de modalités différentes. Dans ce cadre, les approches géométriques ont montré leurs limitations, notamment par la difficulté d'extraire des primitives équivalentes dans des images de natures différentes. En effet, dans le cas des approches géométriques, la qualité de l'estimation de la transformation recherchée est totalement conditionnée par la segmentation des images (détection des primitives géométriques). Les méthodes iconiques les ont peu à peu supplantées en recalage multimodal. Les travaux de Woods [Woods et al., 1993] ont permis une avancée significative en ce sens avec l'introduction du concept d'histogramme conjoint, qui est devenu un outil très efficace du recalage dans le cadre iconique. L'information mutuelle a ensuite été utilisée [Collignon et al., 1995a, Viola, 1995] et son succès ne s'est jamais démenti par la suite. Nous reviendrons en détail dans la suite sur ces mesures de similarité utilisées en recalage iconique.

Les approches iconiques ont la particularité de ne pas nécessiter de prétraitement pour réduire la masse de données, comme c'est le cas des approches géométriques. Elles sont considérées comme les méthodes de recalage les plus flexibles, puisqu'elles utilisent toute l'information des images au cours du processus de recalage. Un autre avantage des méthodes iconiques est leur caractère automatique : elles ne nécessitent pas de sélection manuelle ou de segmentation, souvent délicate, de primitives géométriques communes aux deux images. Cependant, elles présentent tout de même quelques inconvénients. Tout d'abord les caractéristiques extraites de l'image (en général les niveaux de gris) sont peu informatives par rapport aux primitives géométriques. Ensuite, le temps de traitement est en général supérieur à celui des méthodes iconiques, puisqu'aucun prétraitement ne vient réduire la masse initiale de données disponibles. Le tableau 1.1 résume les principaux avantages et inconvénients des méthodes géométriques et iconiques. L'approche mixte, qui suit, n'a pas été incluse dans ce tableau puisque son but est justement de s'affranchir des problèmes liés aux deux approches précédentes, pour n'en garder que les avantages.

1.3.3 Approche mixte

Les approches mixtes combinent plusieurs types de caractéristiques extraites des images à recaler. Le but est d'améliorer la robustesse du recalage en combinant les avantages liés aux caractéristiques utilisées. Trois cas sont envisagés :

- **la combinaison de primitives géométriques de nature différente.** Le lecteur peut se référer par exemple à [Yuille et al., 1992] pour la combinaison de points et de courbes, ou à [Hsu and Ostuni, 1999] pour la combinaison de courbes et de surfaces.
- **la combinaison de différentes informations issues des niveaux de gris,** comme c'est le cas dans [Pluim et al., 2000] où le gradient de l'image et l'information des niveaux de gris sont utilisés conjointement. Shen et Davatzikos [Shen and Davatzikos, 2002], de leur côté, construisent un vecteur d'attributs pour chaque voxel de l'image. Ce vecteur contient à la fois l'intensité du voxel en question, les moments géométriques invariants caractéristiques du voisinage du voxel, et une information issue de la segmentation en 3 classes de l'image.
- **la combinaison des approches géométriques et iconiques.** Dans ce cas de figure, les primitives géométriques servent généralement à contraindre des primitives iconiques. C'est le cas notamment dans [Johnson and Christensen, 2002, Sorzano and Unser, 2005] pour la contrainte par des amers ponctuels, et [Collins et al., 1996, Cachier et al., 2001, Hellier and Barillot, 2001, Cachier et al., 2003, Hellier and Barillot, 2003] pour la contrainte par des primitives courbes (sillons corticaux).

Dans ce qui suit, nous nous concentrerons sur les méthodes iconiques qui sont au centre de ce travail de thèse. Dans la littérature, des études s'intéressent spécifiquement à la comparaison entre les méthodes iconiques et géométriques pour le recalage d'images médicales [Duncan and Ayache, 2000] ou plus spécifiquement le recalage d'images cérébrales [McLaughlin et al., 2002], montrant la supériorité des approches iconiques sur les géométriques. Citons également une étude publiée par de nombreux chercheurs [West et al., 1997] qui compare une dizaine de méthodes de recalage et prône l'utilisation des méthodes iconiques dans le cadre du recalage multimodal d'images médicales. C'est en partie ce qui a motivé notre choix initial de travailler sur les méthodes iconiques. Les approches mixtes combinant différentes informations issues des niveaux de gris pourront néanmoins être envisagées dans un travail futur, l'information apportée par le gradient des images pouvant par exemple s'avérer intéressante dans le cadre de notre méthode.

1.4 Mesures de similarité en recalage iconique

Le recalage iconique se base sur la comparaison locale des intensités. De ce fait, les mesures de similarité iconiques opèrent directement sur les niveaux de gris des images, utilisant ainsi toute l'information disponible. Ces mesures font une hypothèse sur la relation liant les niveaux de gris des deux images à recaler. En général, la nature de cette relation permet une classification des différentes mesures de similarité, comme nous allons le voir dans ce qui suit. Aussi, le choix du critère de similarité est guidé par la nature des images à recaler. Toutefois, il faut également considérer la complexité numérique de telle ou telle mesure. Une mesure permettant d'appréhender des relations très générales entre les images sera souvent plus coûteuse et plus difficile à optimiser.

De nombreuses mesures de similarité ont été proposées dans la littérature, et quelques articles proposent même une étude comparative de ces critères [Penney et al., 1998], [Roche et al., 1999], [Sarrut, 2000]. Une fois encore, nous ne chercherons pas à être exhaustifs, mais plutôt à guider le lecteur et à donner les éléments de référence nécessaires au positionnement de ce travail de thèse. Les mesures de similarité globales sont généralement calculées à partir d'un ensemble de couples d'intensités pris dans l'image de référence et l'image flottante. Avant d'expliquer plus en détail ces mesures, il nous apparaît indispensable de se placer dans un cadre probabiliste que nous décrivons maintenant.

1.4.1 Modèle probabiliste des images

Pour évoquer les mesures de similarité en recalage iconique, il est intéressant sinon nécessaire de formaliser le problème en adoptant un modèle probabiliste des images. Tout d'abord, dans ce qui suit, l'intensité de chaque pixel/voxel de l'image de référence (resp. flottante) sera considérée comme une réalisation d'une Variable Aléatoire (VA) notée $x^{(1)}$ (resp. $x^{(2)}$) à valeur dans l'ensemble $D^{(1)}$ des niveaux de gris (resp. dans $D^{(2)}$). Le tirage aléatoire de $x^{(1)}$ (resp. $x^{(2)}$) représentant cette intensité du pixel ν de l'image de référence (resp. flottante) sera alors noté $x^{(1)}(\nu)$ (resp. $x^{(2)}(\nu)$). Notons $\mathbf{x} = (x^{(1)}, x^{(2)})^\top$ (τ pour transposée) le vecteur aléatoire composé des deux VAs $x^{(1)}$ et $x^{(2)}$. Nous supposons dans la suite que les densités de probabilité marginales de $x^{(1)}$ et $x^{(2)}$, notées $p_{x^{(1)}}$ et $p_{x^{(2)}}$, et la densité de probabilité conjointe de \mathbf{x} , notée $p_{\mathbf{x}}$, sont définies. Remarquons qu'un tel modèle probabiliste des images, couramment employé dans la littérature [Roche, 2001], suppose une certaine stationnarité en termes de loi de probabilité d'un pixel à l'autre de l'image.

Nous donnons dans ce qui suit quelques critères de similarité très utilisés mesurant les dépendances qui existent entre les intensités des images à recaler. Nous employons dès à

présent et dans tout ce qui suit le formalisme que nous venons d'adopter afin d'établir les liens entre ces mesures.

1.4.2 Conservation de l'intensité

Les premières mesures iconiques qui sont apparues en recalage d'images médicales considéraient comme idéale la situation où les images sont identiques à une transformation spatiale près. En effet, ces premières mesures reposaient sur l'hypothèse de conservation de l'information photométrique (niveaux de gris identiques dans les 2 images à recaler). Bien entendu, ce postulat s'avère inexact dans quasiment toutes les situations pratiques rencontrées, mais tant que les images restent relativement semblables (comme cela peut être le cas pour les images monomodales par exemple), cette approximation peut être suffisante. Dans ce contexte, les années 70 ont vu l'apparition des premières mesures de similarité se basant sur l'intensité des images. Ainsi, McGillem et Svedlow [McGillem and Svedlow, 1976] ont justifié théoriquement l'utilisation de la **somme des différences au carré** (SDC), ou norme L2 de la différence, lorsque les images sont identiques à une transformation spatiale et à un bruit blanc gaussien additif près [Fitzpatrick et al., 2000]. Cette mesure est toujours largement utilisée comme critère de similarité au sens du maximum de vraisemblance dans le cadre du recalage monomodal intra-sujet bien que le bruit des images médicales soit rarement gaussien. Cette mesure normalisée par le nombre N de pixels peut s'interpréter comme une estimée de l'erreur quadratique moyenne (EQM) entre les deux images :

$$\frac{1}{N}SDC \approx E[(x^{(1)} - x^{(2)})^2] = \oint_{\mathbb{R}^2} (u^{(1)} - u^{(2)})^2 p_{\mathbf{x}}(\mathbf{u}) d\mathbf{u} \quad (1.2)$$

où $E[\cdot]$ désigne l'opérateur de l'espérance mathématique. Les mesures se basant sur cette hypothèse d'identité se sont alors développées au cours des années 70 et 80. Nous pouvons notamment citer la **somme des différences en valeur absolue** (SDA) [Barnea and Silverman, 1972], ou norme L1, qui peut également être interprétée comme une estimée d'une quantité statistique :

$$\frac{1}{N}SDA \approx E[|x^{(1)} - x^{(2)}|] = \oint_{\mathbb{R}^2} |u^{(1)} - u^{(2)}| p_{\mathbf{x}}(\mathbf{u}) d\mathbf{u} \quad (1.3)$$

Plus particulièrement, les mesures SDC et SDA sont définies par :

$$SDC = \sum_{\text{pixels } \boldsymbol{\nu}} (x^{(1)}(\boldsymbol{\nu}) - x^{(2)}(\boldsymbol{\nu}))^2 \quad (1.4)$$

et

$$SDA = \sum_{\text{pixels } \boldsymbol{\nu}} |x^{(1)}(\boldsymbol{\nu}) - x^{(2)}(\boldsymbol{\nu})| \quad (1.5)$$

D'autres critères ont eu également un certain succès. Nous pouvons par exemple citer le nombre de changement de signes dans l'image de différence [Venot et al., 1984], ou la « pattern intensity », qui pourrait se traduire par « motifs d'intensité » [Buzug and Weese, 1998], et qui opère sur la différence des images, faisant apparaître des structures ou motifs d'intérêt.

Mais les limitations de ces mesures sont rapidement apparues, puisque l'hypothèse de départ (conservation de l'intensité dans les images à recaler) est trop forte en pratique. En effet, leur utilisation se limite à la comparaison d'images monomodales intrapatient dans l'étude de l'évolution de pathologies. Mais les besoins de la médecine se diversifiant avec notamment l'émergence de nombreuses modalités d'imagerie et la combinaison possible des informations portées par ces modalités, les chercheurs se sont alors intéressés à des mesures faisant des hypothèses moins fortes sur l'intensité des images à recaler.

1.4.3 Dépendance linéaire ou affine

L'hypothèse de conservation de l'intensité montrant ses limites (dans un cas monomodal comme l'alignement d'images IRM d'une même séquence, cette hypothèse n'est par exemple déjà plus satisfaite), les chercheurs ont tenté de modéliser des relations plus complexes entre les images. Ainsi, l'étape suivante fut de supposer une dépendance linéaire ou affine entre les intensités des images, qui peut s'exprimer par :

$$x^{(1)} = \alpha x^{(2)} + \beta, (\alpha, \beta) \in \mathbb{R}^2 \quad (1.6)$$

où α et β sont des constantes inconnues. Lorsque le paramètre β est nul, on parle de dépendance linéaire entre les intensités, sinon de dépendance affine. Le critère de référence, dans le cas linéaire, est le **coefficient d'inter-corrélation normalisé** (ICN) [Brown, 1992] qui est donné par :

$$ICN = \frac{\mathbb{E}[x^{(1)}x^{(2)}]}{\sqrt{\mathbb{E}[x^{(1)}]\mathbb{E}[x^{(2)}]}} = \frac{\int_{\mathbb{R}^2} u^{(1)}u^{(2)}p_{\mathbf{x}}(\mathbf{u})d\mathbf{u}}{\sqrt{\int_{\mathbb{R}} u^{(1)}p_{x^{(1)}}(u^{(1)})du^{(1)} \int_{\mathbb{R}} u^{(2)}p_{x^{(2)}}(u^{(2)})du^{(2)}}} \quad (1.7)$$

Mais encore une fois, on s'attache ici en général au recalage monomodal, c'est pourquoi nous ne détaillerons pas davantage cette métrique. De plus, même dans le cadre d'un recalage monomodal d'un même patient, l'hypothèse linéaire n'est pas toujours vérifiée. En outre, la relation de dépendance affine est plus intéressante, mais encore insuffisante pour les cas multimodaux auxquels nous nous intéressons au cours de cette thèse. La principale mesure, dans le cadre d'une dépendance affine, est le **coefficient de corrélation** (CC) [Svedlow and Anuta, 1976] :

$$CC = \frac{Cov(x^{(1)}, x^{(2)})}{\sqrt{Var(x^{(1)})Var(x^{(2)})}} = \frac{\mathbb{E}[(x^{(1)} - \mathbb{E}[x^{(1)}])(x^{(2)} - \mathbb{E}[x^{(2)}])]}{\sqrt{\mathbb{E}[(x^{(1)} - \mathbb{E}[x^{(1)}])^2]\mathbb{E}[(x^{(2)} - \mathbb{E}[x^{(2)}])^2]}} \quad (1.8)$$

faisant ainsi apparaître la covariance entre les deux images à recaler et le produit de leurs variances. A l'inverse de la SSD, cette mesure doit être maximisée. Des variantes de cette mesure peuvent être également utilisées [Brown, 1992]. Cette classe de mesures est particulièrement bien adaptée pour le recalage monomodal et notamment lorsque les contrastes sont différents entre les images à recaler. Ce type d'hypothèses permet de rendre compte du changement de contraste ou de luminance entre les images à recaler. On considère alors que seule une transformation de luminance linéaire intervient entre les deux images à recaler, comme cela peut-être le cas entre des images de même modalité. Cette mesure est une référence en recalage monomodal, mais elle reste sensible aux valeurs aberrantes (par exemple des points qui apparaissent dans une image mais pas dans l'autre).

1.4.4 Dépendance fonctionnelle

La dépendance fonctionnelle est encore moins contraignante qu'une dépendance affine, puisqu'on ne présuppose pas de relation causale liant les deux images à recaler. On fait seulement l'hypothèse ici que la connaissance de la distribution des niveaux de gris d'une image permet de prédire parfaitement la seconde. La réciproque n'est pas forcément vraie. Ainsi, on envisage ici la possibilité d'un recalage entre modalités différentes, les informations communes aux deux images, telles que les structures anatomiques, pouvant être suffisantes pour vérifier cette hypothèse. Cependant, il faut au préalable s'assurer que les deux modalités en question donnent accès à des informations anatomiques assez proches, sinon l'hypothèse se révèle vite discutable et trop forte une fois encore.

Dans ce contexte, une véritable avancée est à mettre à l'actif de Woods [Woods et al., 1993] qui propose en 1993 un critère homonyme permettant de recaler des images IRM et TEP du

cerveau. Pour cela, il s'affranchit de l'hypothèse de dépendance affine entre les intensités en partitionnant une des images en ses ensembles d'iso-intensités. Puis il calcule, dans chacun de ces sous-ensembles, un écart-type normalisé traduisant la dispersion en intensité des voxels correspondant dans la seconde image. Le **critère de Woods** est alors obtenu en faisant une somme pondérée de ces écarts-types. Pour l'écrire, nous devons tout d'abord introduire les notations de densités de probabilité ($p_{x^{(1)}|x^{(2)}}$), d'espérances ($m_{x^{(1)}|x^{(2)}}$) et de variances ($\sigma_{x^{(1)}|x^{(2)}}^2$) conditionnelles :

$$\begin{aligned} p_{x^{(1)}|x^{(2)}} &= \frac{p_{\mathbf{x}}}{p_{x^{(2)}}} \\ m_{x^{(1)}|x^{(2)}=u^{(2)}} &= \mathbb{E}[x^{(1)}|x^{(2)}=u^{(2)}] \\ &= \int_{\mathbb{R}} u^{(1)} p_{x^{(1)}|x^{(2)}=u^{(2)}}(u^{(1)}) du^{(1)} \\ \sigma_{x^{(1)}|x^{(2)}=u^{(2)}} &= \text{Var}(x^{(1)}|x^{(2)}=u^{(2)}) \\ &= \int_{\mathbb{R}} (u^{(1)} - m_{x^{(1)}|x^{(2)}=u^{(2)}})^2 p_{x^{(1)}|x^{(2)}=u^{(2)}}(u^{(1)}) du^{(1)} \end{aligned} \quad (1.9)$$

Ainsi, le critère de Woods s'écrit :

$$\text{Woods}(x^{(1)}|x^{(2)}) = \mathbb{E}\left[\frac{\sigma_{x^{(1)}|x^{(2)}}}{m_{x^{(1)}|x^{(2)}}}\right] = \int_{\mathbb{R}} \frac{\sigma_{x^{(1)}|x^{(2)}=u^{(2)}}}{m_{x^{(1)}|x^{(2)}=u^{(2)}}} p_{x^{(2)}}(u^{(2)}) du^{(2)} \quad (1.10)$$

Ce critère permet de mesurer la moyenne de la déviation standard normalisée, rapport entre l'écart type et la moyenne. Nous pouvons remarquer que cette mesure est asymétrique, au sens où les deux images ne jouent pas le même rôle dans cette mesure. Ainsi, il est primordial de déterminer à priori l'image qui décrit potentiellement le mieux l'autre. Dans les études de Woods, l'image IRM est prise comme référence, alors que l'image TEP sert d'image flottante. Il faut également noter que bien qu'iconique, cette mesure est utilisée par Woods avec une étape de segmentation préalable afin de supprimer les voxels n'appartenant pas au cerveau. Finalement, différentes variantes de cette mesure ont été proposées dans la littérature [Alpert et al., 1996, Nikou et al., 1998a]. Le **critère d'Alpert** (CA), par exemple, s'exprime par :

$$CA = \mathbb{E}\left[\left(\frac{m_{x^{(2)}}}{m_{x^{(1)}}}x^{(1)} - x^{(2)}\right)^2\right] = \oint_{\mathbb{R}^2} \left(\frac{m_{x^{(2)}}}{m_{x^{(1)}}}u^{(1)} - u^{(2)}\right)^2 p_{\mathbf{x}}(\mathbf{u}) d\mathbf{u} \quad (1.11)$$

où $m_{x^{(1)}} = \mathbb{E}[x^{(1)}]$ et $m_{x^{(2)}} = \mathbb{E}[x^{(2)}]$.

L'autre grande mesure qui fait l'hypothèse d'une dépendance fonctionnelle est le **rapport de corrélation** (RC), une mesure statistique liée à la notion de variance qui a été proposée dans le cadre du recalage multimodalités dans [Roche et al., 1998]. Elle s'écrit :

$$RC^2(x^{(1)}|x^{(2)}) = 1 - \frac{1}{\sigma_{x^{(1)}}^2} \mathbb{E}[\sigma_{x^{(1)}|x^{(2)}}^2] = 1 - \frac{1}{\sigma_{x^{(1)}}^2} \int_{\mathbb{R}} \sigma_{x^{(1)}|x^{(2)}=u^{(2)}}^2 p_{x^{(2)}}(u^{(2)}) du^{(2)} \quad (1.12)$$

où $\sigma_{x^{(1)}}^2 = \text{Var}(x^{(1)})$. Premier anachronisme dans cette présentation, cette mesure qui s'appuie essentiellement sur la variance, a été développée pour contourner les limites de l'information mutuelle que nous développerons au chapitre 2, une mesure qui repose sur des hypothèses

faibles concernant la relation entre les intensités des images. Une fois de plus, cette mesure est asymétrique et s'appuie sur les statistiques des images que sont l'espérance et la variance. L'estimation de ces mesures statistiques est un problème à part entière et le reproche qui a souvent été fait à de telles métriques est qu'en présence de données aberrantes, les estimateurs de l'espérance et de la variance peuvent conduire à un critère imprécis. Pour palier ce problème, des estimateurs robustes ont été proposés, notamment dans [Nikou et al., 1998a, Nikou et al., 1998b]. Cette mesure a été présentée dans [Roche et al., 1998] pour le recalage multimodal IRM/TEP/ scanner. Elle permet donc de s'affranchir en partie du problème cité précédemment concernant les informations anatomiques différentes apportées par les deux types de modalités d'imagerie. Intéressons nous maintenant à la classe de mesures liées à la notion d'entropie, qui fait le moins d'hypothèses quant aux relations unissant les intensités dans les deux images à recalcr.

1.4.5 Dépendance statistique

En ne considérant qu'une relation statistique entre les images, un pas supplémentaire fut franchi vers le recalage multimodal. La notion de dépendance statistique étant plus générale que la notion de dépendance fonctionnelle ou affine, les mesures de cette classe sont celles qui reposent sur les hypothèses les plus faibles. Ceci a été rendu possible grâce au concept **d'histogramme conjoint** introduit par [Hill et al., 1994] visant à estimer la densité de probabilité conjointe p_x des deux images à recalcr. Hill s'est en effet rendu compte que quand deux images sont mal recalées, leur histogramme conjoint tend à se disperser. Ainsi, une mesure prenant en compte la dispersion de l'histogramme conjoint permet de recalcr deux images. Plutôt que d'utiliser seulement la densité de probabilité conjointe p_x , les chercheurs se sont alors tournés vers la théorie de l'information, et le cadre probabiliste de Shannon [Shannon, 1948] pour proposer de nouvelles mesures. Parmi celles-ci, les plus remarquables sont sans doute **l'entropie conjointe** [Collignon et al., 1995b, Studholme and Hawkes, 1995], **l'Information Mutuelle (IM)** [Collignon et al., 1995a, Viola, 1995] et **l'Information Mutuelle Normalisée (IMN)** [Studholme et al., 1999]. La distance de Hellinger, expliquée en détail dans [Swain and Ballard, 1991], basée sur le **coefficient de Bhattacharya**, fait également partie de ces mesures efficaces pour le recalage d'images statistiquement dépendantes [Rousseau, 2003].

Nous ne nous étendons pas ici sur leurs expressions car nous présentons ces mesures issues de la théorie de l'information dans le chapitre suivant, puisqu'elles sont le point de départ de ce travail de thèse. En effet, notre contribution principale a été de construire un nouveau critère de similarité basé sur une approximation de ces mesures probabilistes. Pour ce faire, nous avons exploité un développement des densités de probabilité des images au voisinage des gaussiennes les plus proches faisant intervenir les statistiques d'ordre supérieur. Ce développement sera présenté en détail dans le chapitre 3.

Intéressons nous maintenant aux modèles de déformation qui conditionnent la nature des transformations possibles des images à recalcr.

1.5 Modèles de déformation

Un choix crucial à réaliser avant de s'attaquer à un problème de recalage d'images médicales est le modèle de déformation ou de transformation géométrique à prendre en compte. Celui-ci va définir la façon dont l'image flottante $x^{(2)}(\nu)$ sera déformée pour la faire correspondre au mieux à l'image de référence $x^{(1)}(\nu)$. Ce choix est généralement orienté par trois contraintes :

- **les structures à recalcr.** Le modèle de déformation ne sera pas le même en fonction des structures anatomiques à recalcr. Pour recalcr des structures osseuses par exemple, le modèle de déformation sera différent de celui employé pour recalcr les structures molles telles que le coeur. La déformation des os étant globalement rigide entre 2 images, un modèle simple (rigide ou affine) de déformation est souvent suffisant. Si l'on veut s'intéresser au recalcr d'images cardiaques au contraire, le caractère globalement non-rigide des déformations de l'organe doit nécessairement être pris en compte.
- **la finalité clinique.** Un modèle de déformation affine peut être suffisant pour une mesure d'évolution en recalcr monomodal, tandis que dans le cadre d'un recalcr multimodal, il sera souvent nécessaire de faire appel à des modèles plus complexes.
- **le degré de réalisme du modèle.** Plus un modèle de déformation est réaliste, c'est à dire qu'il permet des transformations hautement non-rigides, plus son nombre de degrés de liberté (DDL) est en général important. Les DDL représentent la dimension de l'espace de recherche, c'est-à-dire le nombre de paramètres à trouver. Pendant le processus d'optimisation, chaque paramètre de transformation doit être identifié. Il apparaît donc évident que plus un modèle est complexe, plus son optimisation s'avèrera délicate et coûteuse en temps de calcul. De plus, pour un modèle à grands nombre de DDL se pose nécessairement le problème d'unicité de la solution : si la fonction de coût ou mesure de similarité choisie n'est pas assez régulière dans l'espace des transformations, plusieurs minimas locaux peuvent apparaître.

Dans cette section, nous présentons les modèles présents dans la littérature du recalcr, en partant des plus simples qui ne comportent que quelques DDL, pour aller jusqu'aux plus compliqués qui peuvent comporter plusieurs milliers de DDL (plus de 12000 en ce qui nous concerne, comme nous le présenterons dans le chapitre 4). Nous verrons que notre choix s'est orienté vers les déformations de formes libres (ou FFD pour Free Form Deformations) basées sur les B-Splines [Rueckert et al., 1999], qui représentent un compromis réaliste pour notre application et compte tenu des contraintes précédemment citées.

Quel que soit le modèle de déformation choisi, les transformations peuvent être globales, c'est à dire appliquées de la même manière (c'est à dire avec le même jeu de paramètres) à tous les points de l'image, ou locales, l'image étant alors découpée en régions possédant leur propre transformation. En général, les transformations comportant un faible nombre de DDL (transformations rigides, affines et projectives) sont appliquées globalement, tandis que les transformations non-rigides le sont localement. Une illustration de toutes ces transformations est donnée sur la figure 1.2. De plus, une procédure de recalcr est souvent initialisée grâce à un recalcr rigide ou affine global pour corriger les grandes déformations ou les différences inhérentes aux modalités d'imagerie, et elle se poursuit par un recalcr non-rigide local pour capturer plus finement les déformations. Ainsi, la transformation finale s'écrit :

$$\mathbf{T}(\boldsymbol{\nu}) = \mathbf{T}_{global}(\boldsymbol{\nu}) + \mathbf{T}_{local}(\boldsymbol{\nu}) \quad (1.13)$$

Cet état de l'art sera encore une fois succinct et orienté vers notre problématique. Le lecteur désirant approfondir le sujet peut se référer notamment à [Modersitzki, 2004], qui traite essentiellement des méthodes paramétriques qui seront évoquées dans la suite. Des états de l'art sont également disponibles dans [Elsen et al., 1993, Lester and Arridge, 1999, Rohr, 2000]. Enfin, des études comparatives de plusieurs algorithmes de référence ont été menées par de nombreux chercheurs [West et al., 1997, Klein et al., 2009]. Dans cette dernière étude qui s'intéresse au recalcr non-rigide d'images cérébrales IRM, l'algorithme basé sur les B-Splines est donné comme l'un des plus performants à l'heure actuelle. Ceci a conforté notre choix pour ce modèle de déformation.

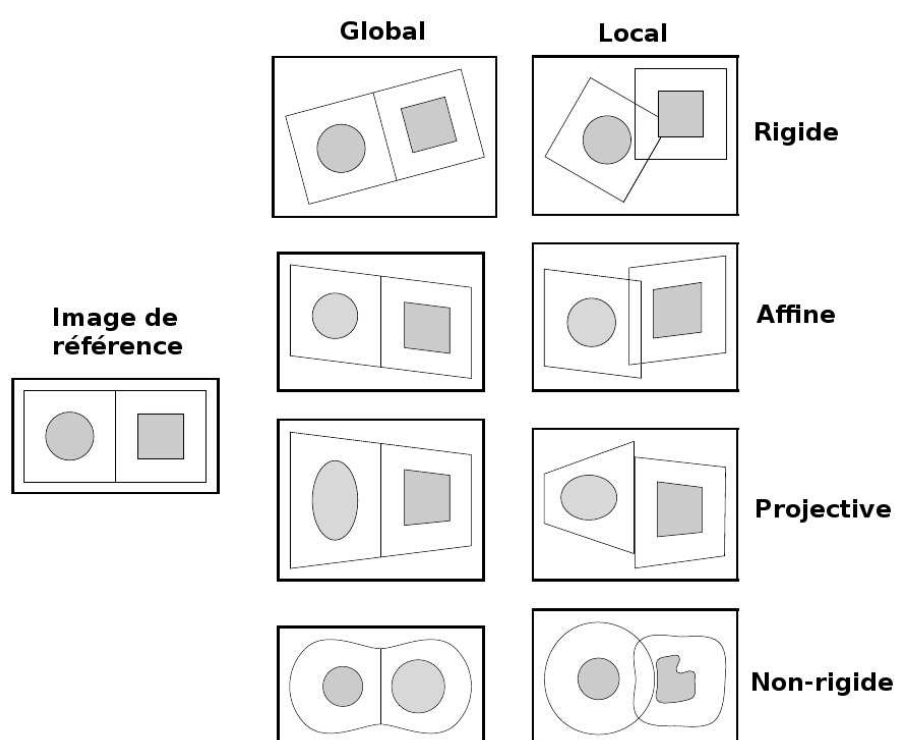


FIGURE 1.2 – Représentations des différents modèles de déformation (d'après [Maintz and Viergever, 1998]).

1.5.1 Transformations rigides

Les transformations rigides ne prennent en compte que des rotations et translations de l'image, et préservent donc les angles et les longueurs (isométrie). Elles appartiennent en général au plus petit espace de recherche utilisé en recalage, puisqu'elles ne comportent que 3 DDL (1 rotation et 2 translations) pour des images 2D ou 6 DDL (3 rotations et 3 translations) pour des volumes d'images. Ces transformations ne conviennent en général que pour des approches monomodales, quand par exemple on cherche à compenser la différence de positionnement d'un patient par rapport aux capteurs sans tenir compte d'éventuelles distortions des tissus imagés. Elles peuvent également être choisies comme étape préliminaire avant un recalage non-rigide. Dans ce qui suit, nous allons présenter les différents types de transformations sous forme matricielle et pour des images tri-dimensionnelles. Ainsi, la transformation rigide recalant une image flottante $x^{(2)}(\boldsymbol{\nu})$ sur une image de référence $x^{(1)}(\boldsymbol{\nu})$, c'est à dire permettant d'obtenir l'égalité suivante :

$$\forall \boldsymbol{\nu}, x^{(1)}(\boldsymbol{\nu}) = x^{(2)}(\mathbf{M}_{\mathbf{R}}(\boldsymbol{\nu}, 1)^{\top}) \quad (1.14)$$

peut être définie par $\mathbf{M}_{\mathbf{R}}$, une matrice de transformation rigide obtenue comme le produit matriciel de matrices de rotation et de translation. La paramétrisation de ces matrices est donnée en annexe A.

1.5.2 Transformations affines

Ces classes de transformations sont utilisées pour des recalages de structures relativement rigides, ou pour initialiser un recalage non-rigide. C'est ce second cas qui nous intéresse. En effet, nous avons vu que les transformations affines ou rigides sont en général appliquées globalement à l'image, tandis que les transformations non-rigides que nous allons présenter en détail dans la suite s'appliquent localement. Cette classification n'est pas toujours exacte, et il faut tout de même noter l'intérêt de l'utilisation des transformations affines par morceaux dans certains pour un recalage inter-sujets [Talairach and Tournoux, 1988]. Mais dans la suite, nous ne considérons que les transformations globales affines. Il est souvent intéressant de commencer une procédure de recalage non-rigide par un recalage considérant ce type de transformation. Cela permet de compenser les grandes déformations dues par exemple au positionnement du patient sur la table d'examen, ou tout simplement à la différence entre les modalités d'imagerie (correction du facteur d'échelle). Pour obtenir une transformation affine, il faut ajouter au modèle rigide les facteurs d'échelle et de cisaillement. En 3D, on compte 3 facteurs d'échelle (1 dans chaque direction) et 6 coefficients de cisaillement. Ceci porte à 15 le nombre de DDL en 3D (et à 7 en 2D). Les transformations sont appliquées matriciellement suivant l'équation (1.14), à ceci près que la matrice de transformation rigide $\mathbf{M}_{\mathbf{R}}$ est remplacée par une matrice affine $\mathbf{M}_{\mathbf{A}}$. Plus de détails quant à la paramétrisation de cette matrice sont donnés une fois encore en annexe A.

Nous n'évoquerons pas ici les transformations projectives qui ne sont que rarement utilisées en recalage d'images médicales. Quand elles le sont, c'est en général pour des opérations de recalage 2D/3D.

Grâce à ces 2 modèles de déformations relativement simples, il est déjà possible de représenter un grand nombre de déformations. Cependant, notre sujet initial étant le recalage iconique d'organes, et étant donné le caractère non-rigide de ces déformations, nous avons dû nous doter d'un modèle de transformation plus réaliste. Celui-ci doit en effet être capable de modéliser convenablement les déformations complexes d'un ensemble de tissus ou les distortions

géométriques inhérentes à certaines modalités d'acquisition. Dans la suite, nous présentons les principaux modèles de déformation non-rigide utilisés en recalage iconique d'images médicales, et nous nous concentrerons sur l'approche que nous avons sélectionnée.

1.5.3 Transformations non-rigides

Contrairement aux modèles rigides et affines, les déformations non-rigides sont appliquées localement, c'est à dire que la transformation appliquée en un point peut effectivement être différente de celle appliquée à ses voisins. Elles correspondent donc au terme $\mathbf{T}_{local}(\boldsymbol{\nu})$ de l'équation (1.13) et interviennent dans notre méthode après un premier recalage affine. L'amplitude des déformations recherchées est généralement plus faible et localisée. Un nombre important de modèles de déformation non-rigide ont été proposés dans la littérature du recalage. Parmi ceux-ci, nous ferons la distinction entre :

- **les approches non-paramétriques**, dans lesquelles la transformation est définie en chaque pixel de l'image. Ces modèles nécessitent l'utilisation d'un terme de régularisation pour contraindre la solution, car le modèle de déformation est totalement libre.
- **les approches paramétriques**, qui sont basées sur l'interpolation par des polynômes (généralement continus par morceau) d'un champ de déformation, en utilisant un ensemble de Points De Contrôle (PDC) placés dans le domaine de l'image. Ces approches, dont notre modèle de déformation est issu, sont contraintes implicitement par les polynômes utilisés pour modéliser le champ de déformation. Bien que plus coûteuses que les approches non-paramétriques, on les préfère généralement pour des applications de recalage multimodal. Il faut noter que les déformations rigides et affines que nous venons de présenter peuvent également être considérées comme paramétriques, puisque contraintes par les coefficients des matrices \mathbf{M}_R et \mathbf{M}_A de transformation qui imposent un nombre fini de DDL.

1.5.3.1 Approches non-paramétriques

Ce premier type de méthode considère des transformations libres, non contraintes par un espace de faible dimension. Cette approche est définie par la classe de transformations la plus large possible, puisque chaque voxel de l'image possède un vecteur de déplacement indépendant. Dans ce cas, il a été montré que le problème d'optimisation est mal posé au sens d'Hadamard [Hadamard, 1923]. On considère qu'un problème est mathématiquement bien posé s'il existe une solution, que cette solution est unique et qu'elle dépend de manière continue des données. Or, par cette approche, la mesure de similarité ne dépend de la transformation que via un nombre fini de points correspondant aux voxels de l'image de référence $x^{(1)}(\boldsymbol{\nu})$. En effet, deux transformations peuvent être différentes tout en coïncidant sur la grille de l'image, et on voit alors que la seule mesure de similarité ne suffit pas à les départager. Il faut alors régulariser le problème d'optimisation en suivant la théorie de l'approximation de Tikhonov. Cela revient à ajouter à la mesure de similarité un terme de régularisation indépendant des images et qui contraint l'espace des transformations possibles, pour le restreindre par exemple à des difféomorphismes. Le problème de recalage peut alors se formaliser par :

$$\mathbf{T}_o = \arg \min_{\mathbf{T}} [S(f(x^{(1)}(\boldsymbol{\nu})), f(x^{(2)}(\mathbf{T}(\boldsymbol{\nu})))) + \lambda R(\mathbf{T})] \quad (1.15)$$

où $S(f(x^{(1)}(\boldsymbol{\nu})), f(x^{(2)}(\mathbf{T}(\boldsymbol{\nu}))))$ représente la mesure de similarité, $R(\mathbf{T})$ est le terme de régularisation et λ est un scalaire qui contrôle l'influence relative des deux termes. C'est sur le terme de régularisation que nous allons porter notre attention, car nombre de travaux se sont intéressés à sa construction. La régularisation du problème de recalage est nécessaire pour

contraindre l'espace des solutions à ne considérer que des champs de déformations « physiquement acceptables » (une structure osseuse doit rester rigide par exemple). Elle s'exprime en contraignant la déformation par des forces internes dépendant des propriétés comportementales des transformations considérées. Nous présentons ici les principales énergies de régularisation utilisées en recalage d'images médicales. Il faut noter que peu d'entre-elles s'utilisent avec l'information mutuelle comme mesure de similarité, c'est pourquoi nous ne les présenterons que succinctement.

Flux optique La méthode de flux optique est une technique de recalage basée sur le flot optique, qui est équivalent à l'équation de mouvement des écoulements incompressibles en mécanique des fluides. Cette méthode faisant l'hypothèse d'invariance d'intensité entre les images, elle est réservée en général au recalage monomodal. Les inventeurs de cette approche sont Horn et Schunk [Horn and Schunck, 1981]. Leur terme de régularisation est issu de la théorie de l'estimation de mouvement, et plus particulièrement du flux optique, correspondant au modèle de membrane.

Modèle élastique Le modèle physique élastique considère l'image flottante comme un solide élastique linéaire [Broit, 1981, Bajcsy and Broit, 1982] et la déforme en utilisant des forces issues de la mesure de similarité choisie. Deux forces s'opposent alors : la force élastique interne qui s'oppose à la force externe de la mesure de similarité. L'image flottante est alors déformée jusqu'à ce que les 2 forces atteignent un équilibre. L'hypothèse d'élasticité linéaire n'est valable que pour des petites déformations [Bajcsy and Kovacic, 1989], il est donc difficile de retrouver de grandes transformations grâce à cette régularisation. Ce terme de régularisation, lié à la physique des objets du recalage, est donné par le potentiel élastique du champ de déplacement. Ce potentiel est régi par l'équation aux dérivées partielles élastique linéaire de Navier :

$$\mu_l \nabla^2 \mathbf{u}(\boldsymbol{\nu}) + (\lambda_l + \mu_l) \nabla(\nabla^\top \mathbf{u}(\boldsymbol{\nu})) + \mathbf{b}(\boldsymbol{\nu}) = 0 \quad (1.16)$$

où λ_l et μ_l désignent les constantes de Lamé qui quantifient les propriétés élastiques du modèle, $\mathbf{u}(\boldsymbol{\nu})$ représente le champ de déplacement et $\mathbf{b}(\boldsymbol{\nu})$ représente les forces extérieures subies par le matériau. Dans notre cadre, $\mathbf{b}(\boldsymbol{\nu})$ représente la force visant à mettre en correspondance les caractéristiques communes aux deux images.

Modèle fluide Le modèle fluide [Christensen et al., 1996] permet de rendre compte de déformations importantes et très localisées car il est d'une grande flexibilité. La contrepartie à cette flexibilité est le risque d'aboutir à un mauvais recalage, généralement à cause de la croissance d'une région qui aura remplacé la distortion ou le déplacement d'une autre [Lester and Arridge, 1999]. Les équations aux dérivées partielles qui régissent ce modèle sont les mêmes que celles du modèle élastique (équation (1.16)), à ceci près qu'elles agissent sur la vitesse et non plus sur le déplacement (équations de Navier-Stokes). Ces méthodes sont coûteuses en temps de calcul.

1.5.3.2 Approches paramétriques

L'approche paramétrique est dite auto-contrainte, car la régularisation se fait directement grâce au modèle de déformation. Ces approches sont paramétrées par des PDC répartis régulièrement ou irrégulièrement sur l'image de référence. Chacun des ces PDC est associé à un vecteur de position représentant les paramètres du PDC. La transformation globale est alors obtenue en extrapolant la transformation en dehors des PDC par différentes méthodes. Ces approches possèdent une forme de régularisation intrinsèque à la transformation, due au

modèle utilisé pour exprimer le champ de déformation. Ce champ est modélisé par une somme de fonctions de base $\beta(\nu)$. Ces fonctions peuvent avoir un support global, c'est à dire que leur influence s'étendra sur toute l'image. C'est le cas notamment des fonctions à base radiales (RBF), des fonctions plaques minces (TPS pour Thin Plate Splines) [Meyer et al., 1997] qui sont un cas particulier des RBF, mais également des fonctions polynomiales [Horsfield, 1999], gaussiennes [Lau et al., 2001] ou trigonométriques [Ashburner and Friston, 1999]. Plus intéressantes sont les fonctions à support local, dont l'influence ne s'étend que sur une sous-région de l'image. On trouve dans cette famille les B-Splines [Rueckert et al., 1999] ou les fonctions de base à symétrie radiale localisée (localized RBF) [Rohde et al., 2003].

Un terme de régularisation peut être ajouté à la fonction de coût pour pénaliser des déformations indésirables, et ainsi réduire le nombre de minima locaux. Pour les méthodes de recalage non-paramétriques, ce terme de régularisation adapté est essentiel pour assurer la continuité (et donc la différentiabilité) du champ dense de déformation [Fischer and Modersitzki, 2004]. Cependant, dans les approches paramétriques que nous considérons maintenant, le terme de régularisation n'est pas toujours nécessaire. Nous avons fait le choix de ne pas l'utiliser dans notre étude, et nous le justifions ainsi :

- l'utilisation d'un terme de régularisation complexifie l'optimisation et en augmente le coût calculatoire.
- Roche, dans son rapport de thèse [Roche, 2001], montre que l'effet de la régularisation consiste à se ramener à une forme paramétrique comme définie par l'équation (1.19), avec deux différences. D'une part, dans les approches paramétriques dans lesquelles un terme de régularisation est employé, tous les voxels de l'image de référence sont considérés comme des PDC. D'autre part, la transformation $\mathbf{M}_A(\nu, 1)^T$ n'est pas forcément affine.
- une étude conjointe [Klein et al., 2009] menée par des chercheurs experts dans le domaine du recalage non-rigide d'images médicales, visant à comparer quatorze algorithmes de référence, a montré que l'approche que nous avons choisie (un modèle de déformation basé sur les B-Splines sans terme de régularisation) est l'une des plus robustes à l'heure actuelle.

Tout d'abord, nous devons formaliser notre modèle de transformation paramétrique. Il s'apparente aux déformations dites de formes libres (Free Form Deformation ou FFD) qui ont été introduites à la fin des années 80 [Sederberg and Parry, 1986] dans un contexte général de déformation de modèles géométriques solides. Le principe des FFD consiste à plonger le modèle géométrique, ou dans notre cas l'image à déformer dans une grille régulière composée d'un ensemble de PDC. Cette idée a été réemployée dans le contexte du recalage en imagerie médicale à la fin des années 90. [Rueckert et al., 1999] ont été les initiateurs, et cette méthodologie a depuis rencontré beaucoup de succès auprès de la communauté scientifique. On peut notamment citer des applications en imagerie cardiaque [McLeish et al., 2002, Frangi et al., 2004], du sein [Tanner et al., 2002, Schnabel et al., 2003], du thorax [Loeckx, 2006], etc. Dans la suite, nous noterons \mathcal{P}_k le PDC d'indice k où K représente le nombre total de PDC employés pour modéliser la transformation. Le déplacement $\xi^{(k)} = \left(\xi_x^{(k)}, \xi_y^{(k)}, \xi_z^{(k)} \right)^T$ de \mathcal{P}_k entraîne, suivant la base de fonctions choisie, un déplacement sur l'ensemble ou seulement une partie du domaine de définition. La déformation continue de l'image est calculée en interpolant les déplacements discrets des PDC. Dans le cadre du recalage, l'ensemble des déplacements des PDC constitue l'ensemble $\xi = \xi^{(k)}$ des paramètres à estimer. Il y a donc dans ces méthodes $d \times K$ paramètres à estimer pour des images d -dimensionnelles. Le déplacement d'un point ν de l'image s'exprime alors comme une combinaison linéaire d'un terme de déplacement et d'une fonction

d'interpolation dont le support reste à définir :

$$\phi(\boldsymbol{\nu}, \boldsymbol{\xi}) = \sum_{k \in \mathcal{K}(\boldsymbol{\nu})} \boldsymbol{\xi}^{(k)} \beta(\boldsymbol{\nu}) \quad (1.17)$$

où $\mathcal{K}(\boldsymbol{\nu}) \subset K$ est un sous-ensemble de PDC qui influence le déplacement du point $\boldsymbol{\nu}$ et $(\beta(\boldsymbol{\nu}))$ est la base de fonctions choisie. Généralement, on introduit l'hypothèse de séparabilité dans chaque dimension de la base choisie, et l'on a donc :

$$\beta(\boldsymbol{\nu}) = \beta_x(\boldsymbol{\nu}) \beta_y(\boldsymbol{\nu}) \beta_z(\boldsymbol{\nu}) \quad (1.18)$$

On voit ici que le choix des PDC va avoir une influence importante sur le type de transformations utilisé. La distribution et le nombre de PDC détermine l'élasticité de la déformation et le nombre de DDL. Dans le cadre de l'approche iconique basée sur les B-Splines [Rueckert et al., 1999], les PDC sont régulièrement espacés. Une approche multiéchelle, que nous expliquerons en détail dans la section 1.7, permet de contrôler l'amplitude des déformations. Ainsi, les déformations locales nécessitent une densité de PDC (et donc un nombre de DDL) plus important pour pouvoir caractériser des déformations au niveau de détail d'un voxel. Ceci est d'autant plus vrai quand les fonctions utilisées ont un support global. Dans les 2 paragraphes suivants, nous présentons les 2 grandes approches paramétriques utilisées en recalage iconique basé sur l'IM, pour montrer ce qui a motivé notre choix des fonctions B-Splines.

Splines de plaques minces (TPS)

L'utilisation des TPS en recalage d'images médicales fut proposé par Bookstein [Bookstein, 1989] à la fin des années 80. Elles sont appelées ainsi car la spline de plaque mince modélise la forme d'une fine plaque de métal déformée. Les TPS font partie d'une famille plus large de splines fondées sur les RBF. Elles forment la fonction radiale la plus utilisée en recalage d'images, même si d'autres fonctions telles que les multiquadratiques ou les fonctions gaussiennes sont également assez répandues. Dans cette approche, les PDC sont considérés comme des paramètres additionnels de la mesure de similarité et sont repositionnés itérativement après avoir été initialisés à la main. De plus, chaque PDC appartenant à une plaque mince a une influence globale sur la transformation, c'est-à-dire que si sa position est perturbée, tous les points de l'image flottante bougent avec lui. Cela limite la capacité de cette méthode à modéliser des déformations complexes et surtout localisées. De plus, quand le nombre de PDC augmente, le coût calculatoire pour bouger un seul PDC augmente également. [Kim et al., 1996] [Meyer et al., 1997] sont les premiers à avoir utilisé les splines de plaque mince dans le cadre du recalage iconique utilisant l'IM comme mesure de similarité. Mais cette approche ne nous semblait pas totalement adaptée à notre problématique, car les transformations que l'on cherche à caractériser sont localisées et de faible amplitude. Aussi, nous nous sommes intéressés à la deuxième grande famille de bases polynomiales utilisées en recalage paramétrique que sont les B-Splines.

B-Splines

Par opposition aux TPS et à leur support global, les B-Splines ont l'avantage d'avoir un support compact. En effet, les B-Splines sont définies au voisinage de leur PDC. Ainsi, perturber la position d'un PDC affecte la déformation seulement au voisinage de ce PDC. Il devient alors possible de modéliser des déformations très localisées. C'est en partie la compacité de leur support qui a fait le succès des B-Splines pour le recalage d'images médicales [Rueckert et al., 1999, Schnabel et al., 2001, Rohlfing et al., 2003], etc. Outre le fait que

cette compacité permet de décrire des transformations très localisées, c'est également un avantage sur le plan du coût calculatoire, car le déplacement d'un PDC n'influence que le voisinage de ce PDC. Mais les B-Splines possèdent d'autres propriétés intéressantes explicitées en annexe B, comme la dérivabilité ou la relative simplicité de mise en oeuvre. Dans la suite, la fonction d'interpolation β définie dans l'équation (1.17) sera choisie comme une fonction de type B-Spline.

La transformation paramétrée par les B-Splines est représentée par une fonction de transformation d -dimensionnelle avec Kd degrés de liberté. Cette fonction $\phi : \mathbb{R}^d \times \mathbb{R}^{Kd} \mapsto \mathbb{R}^d$ associe à chaque voxel $\nu \in \mathbb{R}^d$ de l'image de référence un voxel $\phi(\nu, \xi) \in \mathbb{R}^d$ de l'image flottante en utilisant le vecteur de paramètres $\xi \in \mathbb{R}^{Kd}$. Pour chaque transformation ϕ ou chaque jeu de paramètres ξ , l'image flottante peut être transformée en l'image de référence. Le contenu de la nouvelle image sera l'intensité des voxels et donc l'information portée par l'image flottante, à la position sur l'image de référence correspondante. Pour chaque voxel de référence de coordonnées ν , le vecteur de transformation local $\phi(\nu, \xi)$ est évalué. Puis, l'intensité correspondante dans l'image flottante est estimée en interpolant l'image flottante en $\phi(\nu, \xi)$. Enfin, cette intensité est insérée dans la nouvelle image à la position ν .

En 3D, la transformation $\phi(\nu, \xi)$ est donnée par :

$$\phi(\nu, \xi) = \sum_{k \in \mathcal{K}(\nu)} \xi^{(k)} \beta^{(n)}(\nu) \quad (1.19)$$

avec $\nu = (\nu_x, \nu_y, \nu_z)^\top$ le pixel considéré. La transformation est gouvernée par le vecteur $\xi^{(k)} = (\xi_x^{(k)}, \xi_y^{(k)}, \xi_z^{(k)})^\top$ défini au PDC \mathcal{P}_k de coordonnées $(x_k, y_k, z_k)^\top$. Pour obtenir le déplacement total à la position ν , les déplacements $\xi^{(k)}$ sont pondérés en utilisant les poids des B-Splines $\beta^{(n)}(\nu) = \beta^{(n)}(\nu_x - x_k) \beta^{(n)}(\nu_y - y_k) \beta^{(n)}(\nu_z - z_k)$. Enfin, (n) représente le degré de la B-Spline utilisée (cf. annexe B).

Cette transformation (1.19) correspond à la partie locale de notre transformation dans l'équation (1.13). Notre modèle de transformation final est donc composé d'une transformation globale affine et d'une transformation locale de type FFD B-Spline. La transformation globale peut donc s'écrire, en un point ν de l'image :

$$T(\nu) = M_A(\nu, 1)^\top + \phi(\nu, \xi) \quad (1.20)$$

Pour illustrer notre propos, nous avons utilisé une image scanner du coeur sur laquelle un pixel (en fait un carré de 5×5 pixels pour qu'il soit visible) a été modifié de manière à le faire apparaître. Puis nous avons plongé cette image dans une grille de déformation munie de PDC (1.3). Toutes les illustrations sont données en 2D par souci de clarté, mais la généralisation aux volumes 3D est quasi immédiate. Tout d'abord la pondération de l'influence des PDC est calculée en utilisant un noyau B-Spline. Ainsi, les PDC influenceront le déplacement d'un pixel donné en fonction de leur éloignement à ce pixel, comme présenté figure 1.4. Ainsi, un pixel éloigné d'un PDC ne sera que peu influencé par le déplacement défini en ce PDC. A contrario, un pixel proche du PDC subira un déplacement plus important. Il est également important de comprendre que seuls les PDC au voisinage d'un pixel agissent sur sa déformation, comme illustré figure 1.5(a). De la même manière, un PDC n'a une influence que sur un voisinage $\mathcal{K}(\nu)$ donné (figure 1.5(b)). L'étendue de ce voisinage est donnée par l'étendue du support de la spline (cf. annexe B). Dans ce qui suit, tout comme dans nos expérimentations, nous considérons qu'un pixel/voxel est contrôlé par ses 9 plus proches PDC (en 2D) et ses 27 plus proches (en 3D). La figure 1.6 montre le caractère local du support des B-Splines : en modifiant un PDC, seule une partie de l'image autour de ce PDC est modifiée.

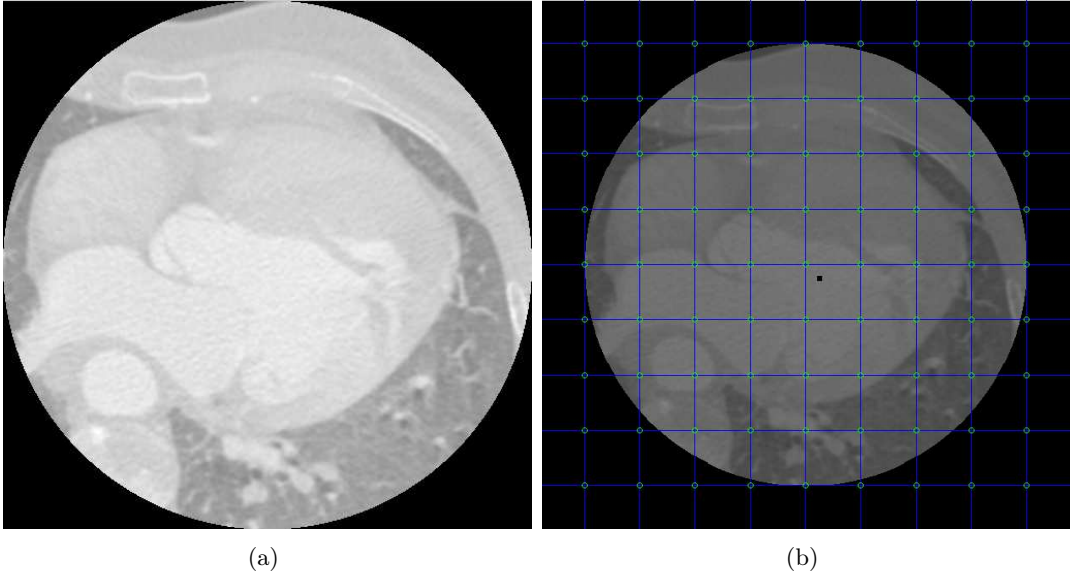


FIGURE 1.3 – Placement de la grille de déformation sur l'image de référence. (a) : image originale. (b) : image modifiée plongée dans la grille de déformation.

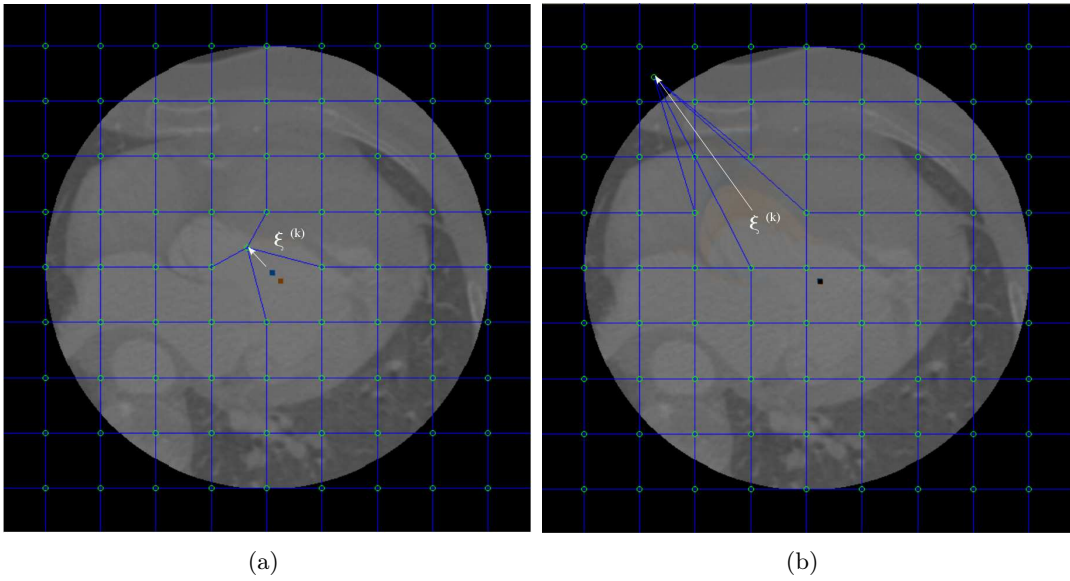


FIGURE 1.4 – Pondération des champs de déplacement des PDC. (a) : Le PDC a légèrement bougé mais il se trouve proche du pixel. Par conséquent, le pixel dans l'image modifiée (en bleu) a beaucoup bougé par rapport au pixel de référence (en rouge). (b) : Cette fois-ci le vecteur de déformation du PDC est très important, mais du fait de son éloignement au pixel test, ce dernier n'est quasiment pas modifié.

1.6 Stratégie d'optimisation

Dans la précédente section, nous avons vu qu'un modèle de déformation correspond en fait à un ensemble de paramètres à estimer. Pour la transformation affine, ces paramètres, contenus

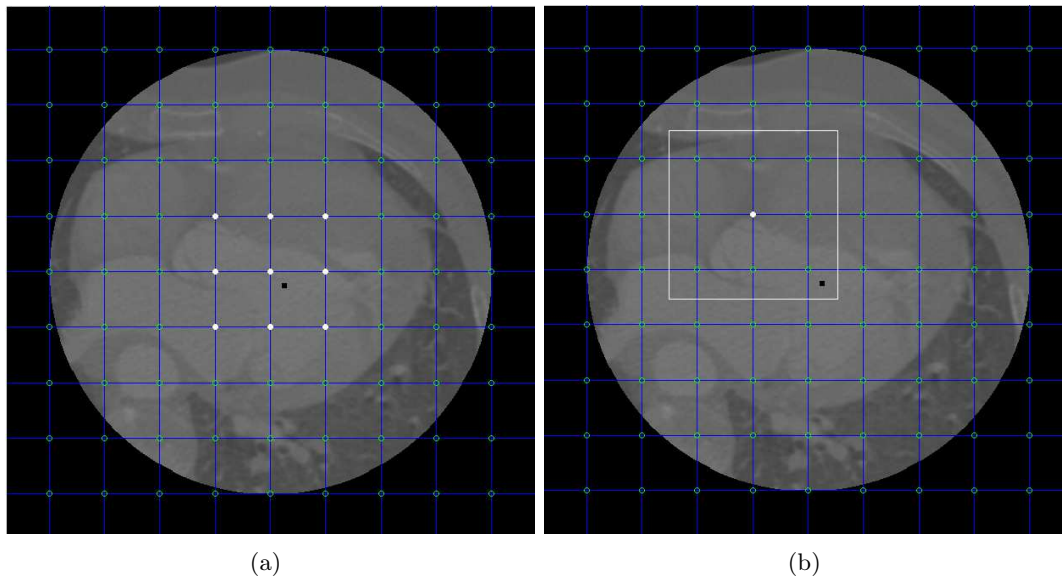


FIGURE 1.5 – Zones d'influence des PDC. (a) : seuls les PDC en blanc participent à la déformation du pixel test. (b) : Un PDC n'a qu'une influence locale, délimitée par le rectangle blanc, sur le champ de déformation.

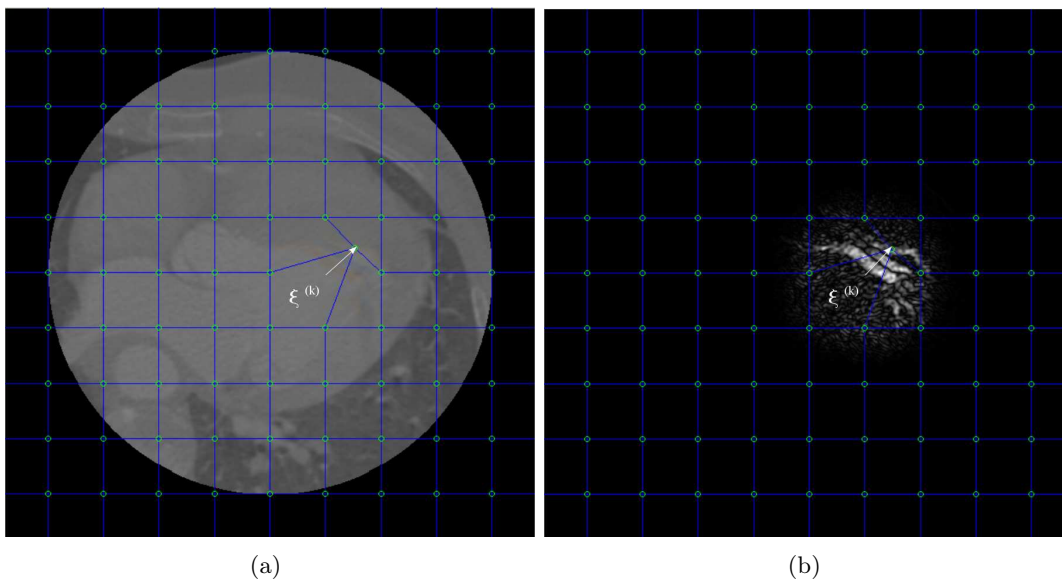


FIGURE 1.6 – Support local des B-Splines. (a) : modification d'un PDC. (b) : différence entre l'image de référence et l'image déformée. Le déplacement du PDC n'a qu'une influence locale sur l'image et qui décroît quand on s'éloigne du PDC. En dehors de ce voisinage d'influence, l'image n'est pas modifiée.

dans la matrice de transformation $\mathbf{M}_{\mathbf{A}}$, sont au nombre de 15 en 3D. Le nombre de paramètres du vecteur ξ de la transformation non-rigide FFD B-Spline est quant à lui variable en fonction du nombre de PDC de la grille de déformation (comme précisé dans la section 1.7), mais l'on compte souvent plusieurs milliers de paramètres à estimer. Une recherche exhaustive dans

l'espace des paramètres est alors totalement inenvisageable, car le temps de traitement serait prohibitif. Il est donc nécessaire d'avoir recours à une méthode d'optimisation, qui va permettre de trouver les paramètres optimaux pour répondre à notre problème de recalage, tout en s'affranchissant de cette recherche exhaustive. Les méthodes d'optimisation stochastiques, qui progressent vers un optimum global, sont peu adaptées à notre problème car le nombre important de paramètres à optimiser rend leur utilisation très coûteuse en temps de calcul. Nous nous sommes donc tournés vers les méthodes déterministes, qui cherchent un optimum local. L'avantage est un coût de traitement réduit. Mais la contrepartie est qu'elles ne convergent pas forcément vers un optimum global, et peuvent rester « bloquées » dans un optimum local. Pour s'en prémunir, il est nécessaire d'initialiser l'algorithme suffisamment près de la solution. Pour y parvenir, l'optimisation est conduite en deux temps : tout d'abord, les paramètres de la transformation affine sont estimés afin de recalculer globalement nos images. Puis, l'optimisation visant à identifier les paramètres de la transformation non-rigide est conduite pour corriger les déformations locales. Le principe reste le même dans les deux cas, c'est pourquoi nous avons choisi de présenter une approche unifiée s'appliquant aux deux situations. Dans la suite, nous noterons $\boldsymbol{\mu}$ le vecteur de paramètres à estimer. L'optimisation consiste à déterminer itérativement le jeu de paramètres $\boldsymbol{\mu}_0$ optimal. L'une des règles d'optimisation itérative les plus populaires est donnée par :

$$\boldsymbol{\mu}_{k+1} = \boldsymbol{\mu}_k + \mathbf{a}_k \boxtimes \mathbf{d}_k, \quad k = 0, 1, 2, \dots \quad (1.21)$$

où \boxtimes est l'opérateur du produit de Hadamard, \mathbf{d}_k représente la « direction de recherche » à l'itération k , et \mathbf{a}_k est le facteur de gain ou pas. Très souvent, $\mathbf{a}_k = a_k \cdot [1, \dots, 1]^T$ avec a_k un facteur de gain scalaire qui contrôle la taille/longueur du pas dans la direction de recherche choisie. Ces deux paramètres sont choisis pour que $\boldsymbol{\mu}_k$ converge vers un minimum de la fonction de coût considérée. La configuration d'un algorithme d'optimisation basé sur la règle (1.21) se ramène donc à la façon de calculer ces 2 paramètres \mathbf{d}_k et a_k . L'article [Klein et al., 2007] s'est intéressé spécifiquement aux méthodes d'optimisation employées dans le cadre du recalage non-rigide utilisant l'IM comme mesure de similarité et les B-Splines comme modèle de déformation, et nous nous appuyerons donc notamment sur cet article pour présenter les approches les plus utilisées pour le calcul des paramètres \mathbf{d}_k et a_k .

1.6.1 Choix du pas

Ce pas va permettre de contrôler la vitesse à laquelle le processus d'optimisation se rapproche de l'optimum. Une vitesse trop lente peut conduire la méthode dans des minima locaux de la fonction de coût, tandis qu'une vitesse trop rapide peut faire passer à côté de l'optimum global. Les méthodes les plus simples consistent à choisir un pas constant, ou un pas qui décroît linéairement en fonction de k . Le problème de ces méthodes est qu'elles sont trop génériques et souvent inadaptées. Elles aboutissent à un processus d'optimisation lent qui peut manquer le minimum global de la fonction de coût. Or, le choix d'un pas acceptable a un impact majeur sur la robustesse et l'efficacité d'une méthode d'optimisation. C'est pourquoi dans nos expérimentations, nous préférons souvent utiliser une procédure à pas adaptatif. Une procédure de recherche minimise à chaque itération la fonction de coût \mathcal{C} suivant la direction de recherche \mathbf{d}_k considérée à cette itération :

$$a_k = \arg \min_a \mathcal{C}(\boldsymbol{\mu}_k + a \mathbf{d}_k) \quad (1.22)$$

Le souci majeur d'une telle procédure est qu'elle engendre un nombre important d'évaluations de la fonction de coût et de ses dérivées. En général, le problème est donc quelque peu

contourné : plutôt que de chercher le pas a_k qui minimise la fonction de coût, on cherche un pas qui donne une réduction suffisante de cette fonction de coût. Deux approches sont proposées ici pour chercher un tel pas

La première se nomme « recherche linéaire inexacte » (inexact line search) et a été donnée par [Moré and Thunberg, 1994]. Elle détermine a_k de sorte que les conditions fortes de Wolfe [Nocedal and Wright, 2000] soient satisfaites :

$$\mathcal{C}(\boldsymbol{\mu}_{k+1}) \leq \mathcal{C}(\boldsymbol{\mu}_k) + c_1 a_k \mathbf{d}_k^T \mathbf{g}(\boldsymbol{\mu}_k) \quad (1.23)$$

$$|\mathbf{d}_k^T \mathbf{g}(\boldsymbol{\mu}_{k+1})| \leq c_2 |\mathbf{d}_k^T \mathbf{g}(\boldsymbol{\mu}_k)| \quad (1.24)$$

où $\mathbf{g}(\boldsymbol{\mu}_k)$ est le gradient de la fonction de coût évalué à la position $\boldsymbol{\mu}_k$, où c_1 et c_2 sont des constantes à définir, en respectant $0 < c_1 < c_2 < 1$. En pratique, c_1 est habituellement choisi très petit, tandis que c_2 est pris beaucoup plus grand. [Potra and Shi, 1995] donnent des valeurs de $c_1 = 10^{-4}$ et $c_2 = 0,9$ pour les méthodes de Newton ou de quasi-Newton (que nous détaillerons dans la suite) et $c_2 = 0,1$ pour la méthode du gradient conjugué. La première condition de Wolfe (1.23), connue sous le nom de condition d'Armijo, impose à a_k de donner une décroissance suffisante de la fonction de coût, tandis que la deuxième (1.24), la condition de courbure ou condition de Goldstein, assure une avancée raisonnable vers un point stationnaire de la fonction de coût, où la dérivée s'annule. Les conditions fortes de Wolfe permettent d'assurer la convergence de la dérivée de la fonction de coût vers zéro.

En pratique, cette procédure de recherche de pas adaptatif se fait à chaque itération de la procédure d'optimisation, et en 2 temps : la phase de « mise en parenthèse » (bracketing phase) et la phase « de coupe » (sectioning phase).

« Bracketing phase »

La première phase consiste à tracer un graphe $C = f(a_k)$ de la fonction de coût en fonction du pas et à trouver 2 bornes α_1 et α_2 , dans ce graphe, entre lesquelles la fonction de coût atteint un minimum. Ceci est fait en respectant les critères de Wolfe (1.23) et (1.24) précédemment cités. Lorsque ce bassin de convergence est obtenu, ses bornes sont enregistrées. Cette phase donne donc deux bornes auxquelles sont associées deux pas ainsi que deux fonctions de coût et leurs dérivées.

« Sectioning phase »

Pendant cette seconde phase, le but est de trouver le minimum de la fonction de coût entre les 2 bornes de l'intervalle calculées précédemment. Pour ce faire, la fonction de coût et sa dérivée sont interpolées à l'aide d'un polynôme cubique, et le minimum de ce polynôme est trouvé. Ceci permet d'obtenir un pas a_k garantissant une décroissance suffisante de la fonction de coût. La notion de décroissance suffisante peut paraître ambiguë. En pratique, des paramètres additionnels permettent lors de ces 2 phases de contrôler cette décroissance. Ce pas a_k est choisi comme pas optimal pour l'itération k de la procédure d'optimisation. Une illustration de cette recherche de pas est donnée figure 1.7.

Le pas a_k de l'équation (1.21) a donc été trouvé par cette procédure de recherche linéaire. Il est encore nécessaire de définir la direction de recherche \mathbf{d}_k de cette même équation. Pour cela, un panel assez large d'algorithmes a été proposé dans la littérature. Nous les détaillons dans ce qui suit.

La seconde méthode qui vise à adapter le pas pour obtenir une décroissance suffisante de la fonction de coût a été proposée récemment [Klein et al., 2009] dans le cadre du recalage d'images et se nomme descente de gradient stochastique à pas adaptatif. Elle se base sur le

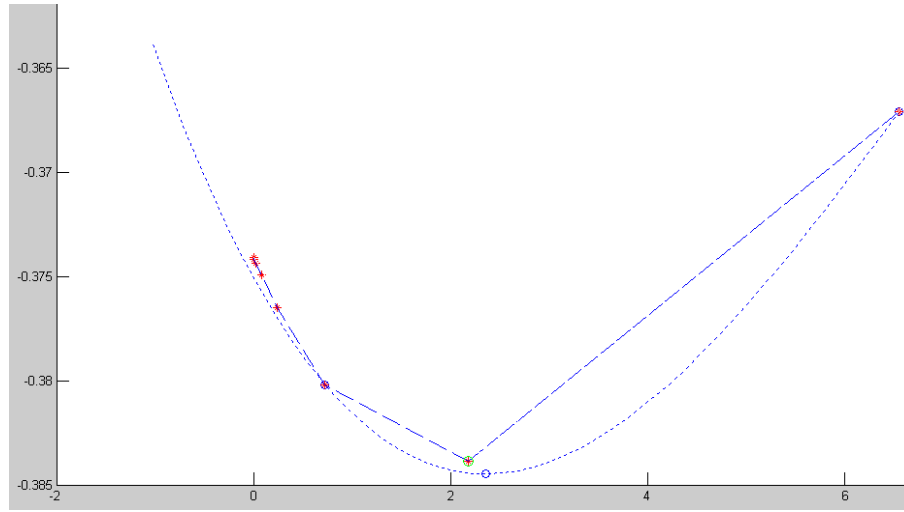


FIGURE 1.7 – Graphe de la fonction de coût en fonction du pas permettant la recherche du pas respectant les conditions fortes de Wolfe : pendant la première phase, la fonction de coût est évaluée pour plusieurs valeurs de a_k (Points étoilés). Une fois qu'un bassin de convergence a été identifié, on cherche le minimum de la fonction de coût (rond bleu) dans ce bassin grâce à un polynôme (en pointillé) interpolant les valeurs de la fonction de coût précédemment calculées. Le pas finalement retenu est celui qui minimise le polynôme sur l'intervalle considéré.

travail théorique mené dans [Plakhov and Cruz, 2004]. Le principal avantage de cet algorithme, outre son efficacité en recalage d'images médicales qui a été démontrée dans [Klein et al., 2009], est que les paramètres « libres » les plus importants de l'algorithme peuvent être estimés automatiquement à partir des caractéristiques des images mises en jeu. En reprenant l'équation (1.21), le pas a_k est dans cette méthode remplacé par :

$$\gamma(t_k) = \frac{a}{(t_k + A)^\alpha}, \quad k = 0, 1, \dots, K \quad (1.25)$$

$$t_{k+1} = \max[t_k + f(-\tilde{\mathbf{g}}_k^\top \tilde{\mathbf{g}}_{k-1}), 0] \quad (1.26)$$

où $\tilde{\mathbf{g}}_k$ est l'approximation de la dérivée de la fonction de coût par rapport aux paramètres de la transformation et f représente une fonction sigmoïde (ou *courbe en S*) définie par :

$$f(x) = f_{\min} + \frac{f_{\max} - f_{\min}}{1 - (f_{\max}/f_{\min})e^{-x/\omega}} \quad (1.27)$$

avec $f_{\max} > 0$, $f_{\min} < 0$ et $\omega > 0$. La force de cette approche est que les paramètres a , f_{\max} , f_{\min} et ω sont estimés automatiquement et une fois pour toutes au début du recalage en fonction des images à recaler. La démonstration, élégante, ne sera pas faite ici car elle est mériterait une présentation détaillée. Le lecteur est donc invité à se référer à l'article original [Klein et al., 2009] pour plus d'informations. Nous verrons au chapitre 4 que cette méthode est très efficace pour déterminer le pas, et plus généralement pour optimiser nos mesures de similarité.

1.6.2 Choix de la direction de recherche

Nous avons choisi de classer les méthodes d'optimisation, qui permettent de choisir une direction de recherche, de la manière suivante :

- les méthodes ne nécessitant pas de calcul du gradient de la fonction de coût en fonction des paramètres de la transformation. Ces méthodes sont les plus simples à mettre en oeuvre, mais elles s'appuient généralement sur une recherche « aléatoire » de la direction de recherche. Elles sont en ce sens généralement peu efficaces quand un grand nombre de paramètres doivent être estimés, ce qui est le cas en recalage non-rigide.
- les méthodes s'appuyant sur les dérivées partielles d'ordre 1 (gradient) de la fonction de coût. Elles sont plus efficaces que les précédentes.
- les méthodes utilisant la matrice hessienne, ou matrice des dérivées partielles de second ordre. En incorporant l'information des dérivées secondes, la direction de recherche s'en trouve améliorée. Habituellement, dans ces méthodes, la matrice hessienne n'est pas calculée analytiquement, car ce serait fastidieux et coûteux en temps de calcul. Nous verrons que des méthodes efficaces permettent d'obtenir une approximation correcte de cette matrice.

Les 2 derniers types de méthodes s'appuient sur le calcul de la dérivée de la fonction de coût par rapport aux paramètres de la transformation. Cette dérivée peut être calculée analytiquement ou numériquement.

1.6.2.1 Méthode ne nécessitant pas de calcul de dérivées

Parmi les méthodes ne nécessitant pas de calcul explicite de la dérivée, la plus communément utilisée est l'**algorithme du simplexe**, qui a été proposé par Nelder et Mead [Nelder, J. A. and Mead, R., 1965]. Il est considéré comme économe vis-à-vis du nombre d'évaluations de la fonction de coût par itération, mais ses propriétés de convergence sont loin d'être satisfaisantes dans des problèmes d'optimisation délicats [Mckinnon, 1999], [Lagarias et al., 1998]. Pour remédier à cette faiblesse, de nombreuses variantes de l'algorithme ont été proposées dans la littérature [Tseng, 1999, Nazareth and Tseng, 2002]. Le but ici est de donner l'idée générale de l'algorithme, ou plutôt de sa variante la plus usitée appelée algorithme d'Amoeba. Pour de plus amples informations sur l'algorithme du simplexe, le lecteur pourra se référer à [Walters et al., 1991].

La démarche générale consiste à construire un simplexe, qui est un polytope (ou polyèdre borné) de $N + 1$ sommets dans un espace à N dimensions (N représente ici le nombre de paramètres ou de DDL de notre modèle de déformation). Dans un plan, par exemple, le simplexe est un triangle. Puis, le sommet « le moins bon » (dans le cas d'une minimisation, c'est le sommet donnant la fonction de coût la plus forte) est remplacé par un nouveau sommet grâce à un ensemble de règles que sont la réflexion, l'étirement, ou la contraction du sommet écarté. Le choix de telle ou telle règle est dicté par la valeur de la fonction de coût en ce nouveau sommet. Cette opération est itérée jusqu'à atteindre le critère de convergence (un nombre d'itérations maximum, ou une différence négligeable entre 2 évaluations successives de la fonction de coût). Pour des espaces à faible dimension, comme c'est le cas des transformations affines, cette méthode peut être intéressante, mais elle est inenvisageable pour estimer des transformations non-rigides. Il faut alors se tourner vers les autres classes de méthodes.

1.6.2.2 Méthodes utilisant les dérivées premières

Les méthodes utilisant le gradient de la fonction de coût sont nombreuses. Nous présentons ici les 2 principales utilisées pour le recalage non-rigide d'images médicales basé sur l'IM, à savoir la descente de gradient et le gradient conjugué non linéaire.

La descente de gradient Comme son nom l'indique, dans un algorithme de descente de gradient visant à minimiser une fonction de coût, la direction de recherche sera prise dans le

sens opposé à celui donné par le gradient de la fonction de coût. Nous pouvons donc réécrire l'équation (1.21) de la façon suivante :

$$\boldsymbol{\mu}_{k+1} = \boldsymbol{\mu}_k - a_k \mathbf{g}(\boldsymbol{\mu}_k) \quad (1.28)$$

où $\mathbf{g}(\boldsymbol{\mu}_k)$ est le gradient de la fonction de coût évalué à la position courante $\boldsymbol{\mu}_k$ par rapport aux paramètres de la transformation. Plusieurs variantes de cet algorithme de descente de gradient ont été proposées, suivant que l'on dispose ou non du gradient analytique de notre fonction de coût. Quand c'est le cas, l'équation (1.28) est employée telle quelle. En revanche, quand cette dérivée n'est pas directement disponible, la dérivée $\mathbf{g}(\boldsymbol{\mu}_k)$ est remplacée par son approximation $\tilde{\mathbf{g}}_k$, ce qui résulte dans le schéma d'optimisation suivant :

$$\boldsymbol{\mu}_{k+1} = \boldsymbol{\mu}_k - a_k \tilde{\mathbf{g}}_k \quad (1.29)$$

Dans cette équation, il reste à choisir la méthode pour approximer le gradient. La plus naturelle consiste à utiliser les différences finies, l'estimation des dérivées étant alors donnée par :

$$[\tilde{\mathbf{g}}_k]_i = \frac{\mathcal{C}(\boldsymbol{\mu}_k + c_k \mathbf{e}_i) - \mathcal{C}(\boldsymbol{\mu}_k - c_k \mathbf{e}_i)}{2c_k} \quad (1.30)$$

où $[\tilde{\mathbf{g}}_k]_i$ représente le i -ème élément de $\tilde{\mathbf{g}}_k$, c_k est un scalaire (idéalement tendant vers 0), et \mathbf{e}_i est un vecteur unitaire constitué de zéros, excepté en l'élément i égal à un. Ici encore, deux possibilités s'offrent : soit la fonction de coût peut être déterminée exactement, soit seule son approximation est disponible. Dans le premier cas, l'équation (1.30) est utilisable directement. Dans le second, on peut par exemple faire appel à la méthode de Kiefer et Wolfowitz [Kiefer and Wolfowitz, 1952] qui suppose que seules des approximations de la fonction de coût sont disponibles. L'approximation de la dérivée (1.30) peut alors se formaliser par :

$$[\tilde{\mathbf{g}}_k]_i = \frac{\tilde{\mathcal{C}}_{ki}^+ - \tilde{\mathcal{C}}_{ki}^-}{2c_k} \quad (1.31)$$

avec :

$$\begin{aligned} \tilde{\mathcal{C}}_{ki}^+ &= \mathcal{C}(\boldsymbol{\mu}_k + c_k \mathbf{e}_i) + \epsilon_{ki}^+ \\ \tilde{\mathcal{C}}_{ki}^- &= \mathcal{C}(\boldsymbol{\mu}_k - c_k \mathbf{e}_i) + \epsilon_{ki}^- \end{aligned} \quad (1.32)$$

où ϵ_{ki}^+ et ϵ_{ki}^- représentent les erreurs d'approximation de la fonction de coût. Ici, l'erreur induite par l'approximation de la dérivée est double : d'une part, une erreur découle du schéma même de la différence finie, de l'autre les termes ϵ de l'équation (1.32) sont une autre source d'erreur à prendre en compte. Aussi, quand cela est possible (c'est à dire quand un accès à la fonction de coût exacte est disponible), nous préférons l'utilisation directe de (1.30). Mais même en utilisant cette équation, la convergence de la solution n'est garantie, d'après [Kushner and Clark, 1978], que si le biais de l'erreur d'approximation tend vers zéro lorsque l'on itère indéfiniment notre algorithme, soit :

$$\mathbb{E}[\tilde{\mathbf{g}}_k] \xrightarrow[k \rightarrow \infty]{} \mathbf{g}(\boldsymbol{\mu}_k) \quad (1.33)$$

En conséquence, les méthodes utilisant une approximation de la dérivée de la fonction de coût ne sont utilisées que lorsque l'évaluation du gradient exact de la fonction de coût est rédhibitoire en termes de temps de calcul, ou lorsque cette évaluation n'est simplement pas disponible.

La gradient conjugué non linéaire Le gradient conjugué non linéaire est une extension du gradient conjugué linéaire développé dans les années 1950 [Hestenes and Stiefel, 1952], et qui permet de minimiser des fonctions non linéaires [Nocedal and Wright, 2000, Dai, 2003]. L'optimisation est toujours menée suivant l'équation (1.21), et la direction de recherche \mathbf{d}_k est définie comme une combinaison linéaire du gradient $\mathbf{g}(\boldsymbol{\mu}_k)$ et de la direction de recherche précédente \mathbf{d}_{k-1} :

$$\mathbf{d}_k = -\mathbf{g}(\boldsymbol{\mu}_k) + \zeta_k \mathbf{d}_{k-1} \quad (1.34)$$

Ici, ζ_k est un scalaire contrôlant l'influence relative du gradient et de la direction de recherche précédente. On trouve dans la littérature [Dai, 2003] plusieurs expressions de ce scalaire, comme par exemple celle de Dai et Yuan [Dai and Yuan, 1999] ou celle de Hestenes et Stiefel [Hestenes and Stiefel, 1952] :

$$\text{Dai-Yuan : } \zeta_k^{DY} = \frac{\mathbf{g}(\boldsymbol{\mu}_k)^\top \mathbf{g}(\boldsymbol{\mu}_k)}{\mathbf{d}_{k-1}^\top (\mathbf{g}(\boldsymbol{\mu}_k) - \mathbf{g}(\boldsymbol{\mu}_{k-1}))} \quad (1.35)$$

$$\text{Hestenes-Stiefel : } \zeta_k^{HS} = \frac{\mathbf{g}(\boldsymbol{\mu}_k)^\top (\mathbf{g}(\boldsymbol{\mu}_k) - \mathbf{g}(\boldsymbol{\mu}_{k-1}))}{\mathbf{d}_{k-1}^\top (\mathbf{g}(\boldsymbol{\mu}_k) - \mathbf{g}(\boldsymbol{\mu}_{k-1}))} \quad (1.36)$$

Le choix de ζ_k aura bien entendu une grande influence sur les propriétés de convergence de la méthode. Nous ne rentrerons pas ici dans les détails, mais le lecteur voulant approfondir le sujet pourra se référer à [Dai, 2003].

Les méthodes de descente de gradient aussi bien que celle de gradient conjugué non linéaire que nous venons de présenter s'associent bien avec une recherche du facteur de gain par procédure de recherche linéaire inexacte, comme présenté dans la section 1.6.1.

La descente de gradient stochastique à pas adaptatif Nous avons présenté cette méthode dans la partie précédente 1.6.1, car son principal apport réside dans la détermination adaptative du pas. Dans cet algorithme [Klein et al., 2009], la direction de recherche est prise dans le sens opposé à celui donné par le gradient de la fonction de coût, tout comme dans une descente de gradient classique. Nous verrons dans le chapitre 4, que cet algorithme est très efficace dans les problèmes de recalage d'images médicales.

1.6.2.3 Méthode utilisant la dérivée seconde

Les méthodes utilisant les dérivées du second ordre sont guidées par l'algorithme de Newton-Raphson, qui est donné par :

$$\boldsymbol{\mu}_{k+1} = \boldsymbol{\mu}_k - [H(\boldsymbol{\mu}_k)]^{-1} \mathbf{g}(\boldsymbol{\mu}_k) \quad (1.37)$$

où $H(\boldsymbol{\mu}_k)$ est la matrice hessienne (ou matrice des dérivées secondes) de la fonction de coût, évaluée en $\boldsymbol{\mu}_k$. L'utilisation de l'information du second ordre donne à l'algorithme d'optimisation de meilleures propriétés de convergence que la descente de gradient. Cependant, le calcul de la matrice hessienne et de son inverse est coûteux en temps de calcul, spécialement pour notre problème de recalage non-rigide où le nombre de paramètres à optimiser est très grand (de l'ordre de plusieurs milliers). Pour contourner cette difficulté, **les méthodes de quasi-Newton** (QN) [Dennis and Moré, 1977, Nocedal and Wright, 2000] s'attaquent au problème en approximant l'inverse de la matrice hessienne par $L_k \approx [H(\boldsymbol{\mu}_k)]^{-1}$. Cette approximation est mise à jour à chaque itération k de l'algorithme d'optimisation. Les dérivées secondes de la fonction de coût ne sont pas nécessaires pour cette mise à jour. Seules les dérivées premières le

sont. Les méthodes QN sont typiquement utilisées en combinaison avec une recherche de pas optimal comme expliqué dans la section 1.6.1. Ainsi, l'algorithme QN est donné par :

$$\boldsymbol{\mu}_{k+1} = \boldsymbol{\mu}_k - a_k L_k \mathbf{g}(\boldsymbol{\mu}_k) \quad (1.38)$$

La construction de L_k est un sujet de recherche à part entière et plusieurs méthodes ont été proposées pour ce faire [Dennis and Moré, 1977, Nocedal and Wright, 2000]. Les plus notables sont sans doute les méthodes « Symmetric-Rank-1 » (SR1), Davidon-Fletcher-Powell (DFP) et Broyden-Fletcher-Goldfarb-Shanno (BFGS). Les expérimentations montrent que la méthode BFGS [Broyden, 1970, Fletcher, 1970, Goldfarb, 1970, Shanno, 1970] est très efficace dans beaucoup d'applications. Elle utilise la règle de mise à jour suivante :

$$L_{k+1} = \left(I - \frac{\mathbf{s}_k \mathbf{y}_k^\top}{\mathbf{s}_k^\top \mathbf{y}_k} \right) L_k \left(I - \frac{\mathbf{y}_k \mathbf{s}_k^\top}{\mathbf{s}_k^\top \mathbf{y}_k} \right) + \frac{\mathbf{s}_k \mathbf{s}_k^\top}{\mathbf{s}_k^\top \mathbf{y}_k} \quad (1.39)$$

avec I la matrice identité, $\mathbf{s}_k = \boldsymbol{\mu}_k - \boldsymbol{\mu}_{k-1}$ et $\mathbf{y}_k = \mathbf{g}(\boldsymbol{\mu}_k) - \mathbf{g}(\boldsymbol{\mu}_{k-1})$.

En pratique cependant, quand le nombre de paramètres devient trop important, la matrice L_k devient très grande, ce qui peut occasionner des problèmes d'allocation de mémoire. C'est pourquoi dans nos expérimentations, nous préférons utiliser une variante connue du BFGS, le Limited memory BFGS (LBFGS) [Nocedal, 1980], qui se prévaut du besoin de stockage de la matrice L_k . Nous utilisons l'approche à deux boucles récursives, initialement proposée par [Liu and Nocedal, 1989].

Nous nous intéressons par la suite à la réduction du temps de calcul que peuvent apporter les approches multirésolutions, qui sont absolument incontournables quand les volumes de données à traiter sont très importants, comme c'est souvent le cas en recalage d'images médicales.

1.7 Multirésolution et approches hiérarchiques

Le calcul du jeu de paramètres de la transformation qui minimise la fonction de coût se fait itérativement pendant l'optimisation. Les itérations peuvent se faire hiérarchiquement du niveau le plus simple au niveau le plus détaillé. La hiérarchisation intervient à deux niveaux dans une procédure de recalage :

- La complexité du modèle de transformation à prendre en compte.
- La complexité des données utilisées.

Cette approche présente deux avantages majeurs. D'une part, elle permet d'accélérer les calculs, ce qui est primordial dans le contexte médical où la finalité est de donner un diagnostic le plus rapidement possible. D'autre part, cela permet également de réduire le risque de converger vers des minima locaux et ainsi d'améliorer la robustesse de la procédure d'optimisation.

Nous présentons dans ce qui suit les choix que nous avons fait concernant cette approche multirésolution.

1.7.1 Complexité du modèle de déformation

Nous envisageons ici une représentation hiérarchisée du modèle de déformation. Celle-ci intervient à 2 niveaux. Comme il a été précisé dans la partie 1.5, nous commençons notre procédure de recalage par un « pré-recalage » affine qui permet de modéliser les grands déplacements (entre 2 modalités d'imagerie par exemple). Puis intervient le recalage non-rigide. Ainsi, ces deux étapes peuvent déjà être considérées comme une approche hiérarchique. En passant d'un modèle de déformation simple à quelques paramètres à un modèle plus élaboré,

la procédure permet de capturer des déformations de plus en plus fines (et donc de se préserver des minima locaux) et de gagner en temps de calcul (la modélisation des grands déplacements par un modèle aussi complexe que les B-Splines est une perte de temps).

La hiérarchisation du modèle de déformation intervient également au niveau du modèle non-rigide lui-même. En effet, ce modèle basé sur les B-Splines est paramétré par un certain nombre de PDC posés sur une grille régulière superposée aux images à recaler. Ces PDC représentent le nombre de DDL du modèle. Comme précisé dans la partie 1.5.3.2, les PDC n'ont qu'une influence locale sur l'image. Ainsi, augmenter leur nombre permet de modéliser des déformations de plus en plus fines et localisées. La procédure de recalage non-rigide tire donc avantage de ce modèle de déformation. Le principe consiste à débiter le recalage avec une grille de PDC espacée. Puis, quand l'optimum est atteint à ce niveau de détail, la grille est affinée en rajoutant des PDC, les paramètres optimaux du niveau précédent étant utilisés pour initialiser cette nouvelle grille. Puis la procédure d'optimisation est relancée avec ces nouveaux PDC qui sont autant de nouveaux paramètres à prendre en compte. Cette approche est souvent qualifiée de pyramide de transformation, comme illustré figure 1.8 en deux dimensions. Le choix du nombre de raffinement de la grille de PDC, ainsi que l'espacement entre ces points, dépend de plusieurs paramètres. Tout d'abord, le but de la pyramide de transformation étant de gagner en temps de calcul, il ne faut pas que le nombre de niveaux de cette pyramide soit trop important. En effet, ceci impliquerait un temps total d'optimisation (la somme des temps d'optimisation à chaque niveau de la pyramide) supérieur au temps de traitement en ne considérant qu'un seul niveau. L'écartement des PDC va quant à lui être dicté par le contenu des images et la complexité des déformations envisagées. Plus les déformations sont localisées et faibles, plus les PDC devront être rapprochés. Cette paramétrisation est une étude à part entière et nous présenterons nos conclusions à ce sujet au chapitre 4.

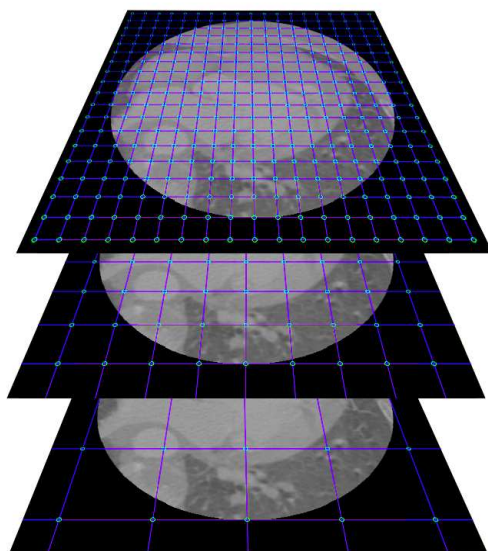


FIGURE 1.8 – Illustration de l'augmentation de la complexité du modèle de déformation par affinage de la grille des PDC B-Splines. Ceci accroît le nombre de DDL et permet ainsi de modéliser des déformations plus faibles et localisées.

1.7.2 Complexité des données

L'autre manière de hiérarchiser la procédure de recalage est d'augmenter progressivement la complexité des données. Les données à traiter sont souvent très volumineuses ($\simeq 512 \times 512 \times 150$ voxels pour des données scanner volumiques en imagerie cardiaque par exemple) et le parcours de ces millions de voxels à chaque itération de la procédure d'optimisation est très coûteux en temps de calcul. L'idée ici est alors d'utiliser un schéma multirésolution telle que la pyramide gaussienne (figure 1.9), permettant de prendre en compte progressivement les détails de l'image. Pour ce faire, des images de taille réduite sont construites par filtrages et sous-échantillonnages successifs. Un filtrage passe-bas gaussien supprime les détails dans l'image, et le sous échantillonnage réduit la taille des images, réduisant par là-même le nombre de voxels à prendre en compte dans la procédure d'optimisation. Comme pour la pyramide de transformation, la question du nombre de niveaux, ou du nombre d'images, dans la pyramide, se pose. Il ne faut pas trop réduire la taille des images, au risque de ne plus avoir assez d'information issue des pixels pour conduire le recalage. Avec la mesure de similarité que nous avons développée et que nous expliciterons dans les chapitres 3, les images sont modélisées par des variables aléatoires (cf. 1.4.1) et les intensités des voxels comme des réalisations de ces variables aléatoires. L'estimée de notre métrique nécessite un nombre « suffisant » d'échantillons des variables aléatoires, nombre que nous étudierons au cours du chapitre 4, de même que celui nécessaire à une bonne estimation de mesures statistiques classiques telles que l'IM.

Nous avons choisi d'utiliser la pyramide gaussienne pour réduire la complexité des données, car elle est relativement simple à mettre en oeuvre et a déjà démontré ses qualités. Néanmoins, d'autres méthodes telles les pyramides d'ondelettes [Deubler and Olivo, 1997] ou les opérations morphologiques peuvent également être utilisées dans ce but. Le lecteur pourra se référer à l'état de l'art [Maes et al., 1999] concernant les méthodes d'optimisation multirésolution et surtout à [Lester and Arridge, 1999] pour une revue bibliographique des différentes approches hiérarchiques et une comparaison des différents types de pyramides (gaussienne, spline, ondelettes).

1.7.3 Approche bi-pyramidale

Dans nos expérimentations, nous avons choisi d'utiliser une approche bi-pyramidale. Ceci consiste à alterner entre les deux pyramides présentées précédemment en suivant le schéma de la figure 1.10. Nous verrons au chapitre 4 les choix que nous avons fait concernant par exemple le nombre de niveaux dans chacune des pyramides durant nos expérimentations.

1.8 Conclusion

Cet état de l'art nous a permis de présenter les problématiques posées et les techniques utilisées en recalage d'images médicales en mettant en évidence à chaque étape les choix réalisés dans ce travail. Comme la notion couvre un champ assez vaste de méthodes, nous nous sommes volontairement focalisés au cours du chapitre sur notre sujet d'intérêt, c'est à dire le recalage non-rigide d'images multimodales. Pour ce faire, nous avons présenté chacune des composantes d'une méthode de recalage (caractéristiques issues des images, mesure de similarité, modèle de déformation, optimisation, etc.) tout en essayant de guider le lecteur vers les choix que nous avons fait en fonction de nos objectifs initiaux. Le but ici n'était pas d'être exhaustif, au regard du nombre conséquent de méthodes proposées dans la littérature. En nous concentrant sur les composantes qui nous ont intéressé au cours de ce travail, nous avons vu que les méthodes iconiques, et plus particulièrement celles utilisant l'IM comme mesure de similarité,

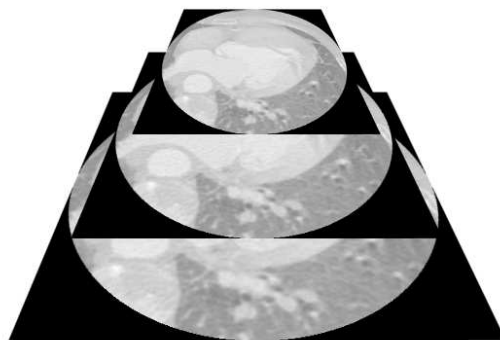


FIGURE 1.9 – Exemple de pyramide gaussienne. Le recalage sera conduit au départ sur l'image la plus petite. Puis les paramètres optimaux à un niveau donné permettent d'initialiser le niveau supérieur.

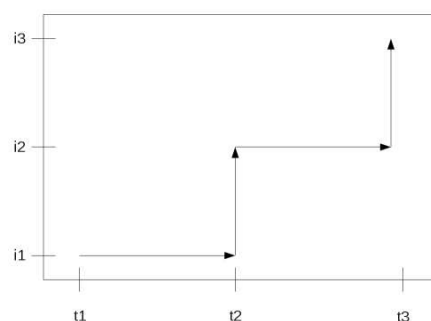
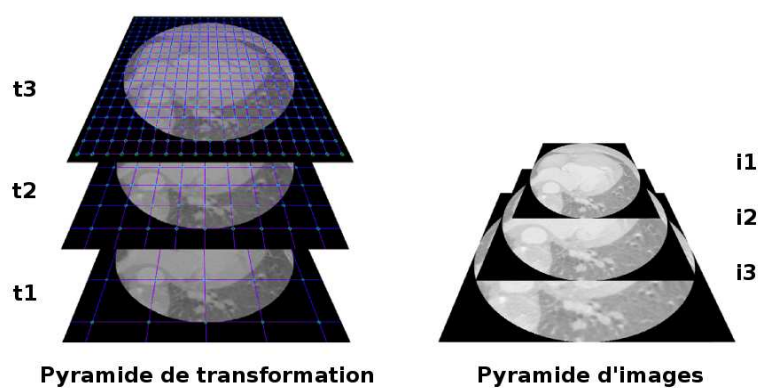


FIGURE 1.10 – Illustration de l'approche bi-pyramidale. Le schéma décrit l'alternance entre les différents niveaux de la pyramide.

sont à l'heure actuelle les plus efficaces pour répondre à notre problème. Le coeur de ce travail de thèse portant sur la proposition d'une nouvelle méthode d'estimation de l'IM, il nous a semblé judicieux d'y consacrer le chapitre qui vient afin de bien mettre en évidence cette mesure statistique et les différentes manières de l'estimer. En effet, l'IM peut être associée à différentes définitions, méthodes et approximations. Le but de ce prochain chapitre est de préciser la démarche qui nous a amené à définir l'IM approximée par une fonction particulière des cumulants d'ordre supérieurs ou égal à deux, qui sera quant à elle présentée en détail dans le chapitre 3.

Bibliographie

- [Acosta et al., 2011] Acosta, O., Simon, A., Monge, F., Commandeur, F., Bassirou, C., Cazoulat, G., de Crevoisier, R., and Haigron, P. (2011). Evaluation of multi-atlas-based segmentation of ct scans in prostate cancer radiotherapy. *IEEE International Symposium on Biomedical Imaging (ISBI)*.
- [Alpert et al., 1996] Alpert, N. M., Berdichevsky, D., Levin, Z., Morris, E. D., and Fischman, A. J. (1996). Improved methods for image registration. *NeuroImage*, 3(1) :10–18.
- [Arad et al., 1994] Arad, N., Dyn, N., Reissfeld, D., and Yeshurun, Y. (1994). Image warping by Radial Basis Functions : application to facial expressions. *Conference on Computer Vision, Graphics and Image Processing*, 56(2) :161–172.
- [Arun et al., 1987] Arun, K. S., Huang, T. S., and Blostein, S. D. (1987). Least-squares fitting of two 3-d point sets. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(5) :698–700.
- [Ashburner and Friston, 1999] Ashburner, J. and Friston, K. (1999). Nonlinear spatial normalization using basis functions. *Human Brain Mapping*, 7(4) :254–66.
- [Atif, 2004] Atif, J. (2004). *Recalage non-rigide multimodal des images radiologiques par information mutuelle quadratique normalisée*. PhD thesis, LIMSI-CNRS Université de Paris XI - Orsay, France.
- [Bajcsy and Broit, 1982] Bajcsy, R. and Broit, C. (1982). Matching of deformed images,. In *Sixth International Conference on Pattern Recognition*, pages 351–353.
- [Bajcsy and Kovacic, 1989] Bajcsy, R. and Kovacic, S. (1989). Multiresolution elastic matching. *Computer Vision, Graphics and Image Processing*, 46 :1–21.
- [Barillot, 1999] Barillot, C. (1999). *Fusion de Données et Imagerie 3D en Médecine*. Habilitation à Diriger des Recherches, Université de Rennes 1 - Institut de Formation Supérieure en Informatique et Communication - Rennes, France.
- [Barnea and Silverman, 1972] Barnea, D. I. and Silverman, H. F. (1972). A class of algorithms for fast digital image registration. *IEEE Transactions on Computers*, 21(2) :179–186.
- [Besl and McKay, 1992] Besl, P. J. and McKay, H. D. (1992). A method for registration of 3-D shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2) :239–256.
- [Bookstein, 1989] Bookstein, F. (1989). Principal warps : Thin-plate splines and the decomposition of deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11 :567–585.
- [Broit, 1981] Broit, C. (1981). *Optimal registration of deformed images*. PhD thesis, University of Pennsylvania, PA, USA.
- [Brown, 1992] Brown, L. G. (1992). A survey of image registration techniques. *ACM Computing Surveys*, 24(4) :325–376.

- [Broyden, 1970] Broyden, C. G. (1970). The convergence of a class of double-rank minimization algorithms. *Journal of the Institute of Mathematics and its Applications*, 6 :76–90.
- [Buzug and Weese, 1998] Buzug, T. M. and Weese, J. (1998). Voxel-based similarity measures for medical image registration in radiological diagnosis and image guided surgery. *Journal of Computing and Information Technology*, 6(2) :165–179.
- [Cachier, 2002] Cachier, P. (2002). *Recalage non rigide d’images médicales volumiques : contributions aux approches iconiques et géométriques*. PhD thesis, Ecole Centrale, Paris, France.
- [Cachier et al., 2003] Cachier, P., Bardinet, E., Dormont, D., Pennec, X., and Ayache, N. (2003). Iconic feature based nonrigid registration : the pasha algorithm. *Computer Vision and Image Understanding*, 89(2-3) :272 – 298.
- [Cachier et al., 2001] Cachier, P., Mangin, J., Pennec, X., Rivière, D., Papadopoulos-Orfanos, D., Régis, J., and Ayache, N. (2001). Multisubject deformable, registration of brain MRI using intensity and geometric features. In Niessen, W. and Viergever, M. A., editors, *Medical Image Computing and Computer-Assisted Intervention MICCAI 2001*, volume 2208, pages 734–742, Utrecht (Netherlands).
- [Cazoulat et al., 2011] Cazoulat, G., Simon, A., De Crevoisier, R., Acosta, O., Lesaunier, M., Louvel, G., Lafond, C., Chajon, E., Leseur, J., and Haigron, P. (2011). Recalage non-rigide pour le cumul de dose en radiothérapie guidée par l’image. *Recherche en Imagerie et Technologie pour la Santé (RITS)*.
- [Chenoune, 2008] Chenoune, Y. (2008). *Estimation des déformations myocardiques par analyse d’images*. PhD thesis, Université Paris Est, Paris, France.
- [Christensen et al., 1996] Christensen, G., Rabbit, R., and Miller, M. (1996). Deformables templates using large deformation kinematics. *IEEE Transactions on Image Processing*, 5 :1435–1447.
- [Collignon et al., 1995a] Collignon, A., Maes, F., Delaere, D., Vandermeulen, D., Suetens, P., and Marchal, G. (1995a). Automated Multi-modality Image Registration based on Information Theory. In *Conference on Information Processing in Medical Imaging*, pages 263–274.
- [Collignon et al., 1995b] Collignon, A., Vandermeulen, D., Suetens, P., and Marchal, G. (1995b). 3D Multi-Modality Medical Image Registration Using Feature Space Clustering. *Lecture Notes in Computer Science*, 905 :195–204.
- [Collins et al., 1996] Collins, D., Le Goualher, G., Venugopal, R., Caramanos, A., Evans, A., and Barillot, C. (1996). Cortical constraints for non-linear cortical registration. In Hohne, K. and Kikinis, R., editors, *Visualization in Biomedical Computing*, volume 1131 of *Lecture Notes in Computer Science*, pages 307–316. Springer Berlin / Heidelberg. 10.1007/BFb0046968.
- [Crum et al., 2004] Crum, W. R., Hartkens, T., and Hill, D. L. G. (2004). Non-rigid image registration : theory and practice. *British Journal of Radiology*, 77(2) :140–153.
- [Dai, 2003] Dai, Y.-H. (2003). A family of hybrid conjugate gradient methods for unconstrained optimization. *Mathematics of Computation*, 72 :1317–1328.
- [Dai and Yuan, 1999] Dai, Y. H. and Yuan, Y. (1999). A nonlinear conjugate gradient method with a strong global convergence property. *SIAM Journal on Optimization*, 10 :177–182.
- [Davis et al., 1997] Davis, M., Khotanzad, A., Flamming, D., and Harms, S. (1997). A physics-based coordinate transformation for 3D image matching. *IEEE Transactions on Medical Imaging*, 16 :317–328.

- [Delhay, 2006] Delhay, B. (2006). *Estimation spatio-temporelle de mouvement et suivi de structures déformables. Application à l'imagerie dynamique du coeur et du thorax*. PhD thesis, INSA Lyon, France.
- [Dennis and Moré, 1977] Dennis, J. E. and Moré, J. J. (1977). Quasi-Newton Methods, Motivation and Theory. *SIAM Review*, 19(1) :46–89.
- [Deubler and Olivo, 1997] Deubler, J. and Olivo, J.-C. (1997). A wavelet-based multiresolution method to automatically register images. *Journal of Mathematical Imaging and Vision*, 7 :199–209.
- [Dréan et al., 2011] Dréan, G., Monge, F., Simon, A., Cazoulat, G., De Crevoisier, R., Haigron, P., and Acosta, O. (2011). Evaluation de méthodes de recalage dense basées sur l'intensité de ct-scans dans le cadre du traitement du cancer de la prostate par radiothérapie guidée par l'aimage. *Recherche en Imagerie et Technologie pour la Santé (RITS)*.
- [Duncan and Ayache, 2000] Duncan, J. S. and Ayache, N. (2000). Medical image analysis : progress over two decades and the challenges ahead. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22 :85–106.
- [Elsen et al., 1993] Elsen, P., Maintz, J., Pol, E., and Viergever, M. (1993). Medical Image Matching - a Review with Classification. *IEEE Engineering in Medicine and Biology*, 12(1) :26–39.
- [Fischer and Modersitzki, 2004] Fischer, B. and Modersitzki, J. (2004). A unified approach to fast image registration and a new curvature based registration technique. *Linear Algebra and its Applications*, 380 :107–124.
- [Fitzpatrick et al., 2000] Fitzpatrick, J., Hill, D., and Maurer Jr, C. (2000). *Image registration ; Handbook of Medical Imaging*, volume 2 of *Medical Image Processing and Analysis*, chapter 8. SPIE Press, Bellingham, Wash, M. Sonka and J.M. Fitzpatrick edition.
- [Fletcher, 1970] Fletcher, R. (1970). A new approach to variable metric algorithms. *The computer journal*, 13 :317–322.
- [Frangi et al., 2004] Frangi, A., Rueckert, D., Schnabel, J., and Niessen, W. (2004). Automatic construction of multiple object three-dimensional statistical shape models : application to cardiac modeling. In *Yearbook of Medical Informatics 2004 : Towards Clinical Bioinformatics*, pages 450–466. Schattauer Verlagsgesellschaft.
- [Garreau et al., 2006] Garreau, M., Simon, A., Boulmier, D., Coatrieux, J.-L., and Breton, H. L. (2006). Surface and motion extraction in cardiac MSCT imaging for the assessment of left ventricular function. *International Journal of Biomedical Imaging*, 2006(Article ID 37607) :Article ID 37607.
- [Gee et al., 1995] Gee, J., Briquer, L., Barillot, C., Haynor, D., and Bajcsy, R. (1995). Bayesian approach to the brain image matching problem. *SPIE Medical Imaging*, 2434(145) :145–156.
- [Gilardi, 1996] Gilardi, M. C. Rizzo, G. (1996). Registration of multi-modal biomedical images of the heart. *The quarterly journal of nuclear medicine*, 40 :142–150.
- [Goldfarb, 1970] Goldfarb, D. (1970). A family of variable metric updates derived by variational means. *Mathematics of Computation*, 24 :23–26.
- [Hadamard, 1923] Hadamard, J. (1923). *[Hadamard, 1923] , Lectures on Cauchy's problem in linear partial differential equations*, New Haven, 1923. New Haven Yale University Press.
- [Hellier and Barillot, 2001] Hellier, P. and Barillot, C. (2001). Cooperation between local and global approaches to register brain images. In *IPMI2001, number 2082 in LNCS*, pages 315–328. Springer-Verlag.

- [Hellier and Barillot, 2003] Hellier, P. and Barillot, C. (2003). Coupling dense and landmark-based approaches for non rigid registration. *IEEE Transactions on Medical Imaging*, 22 :217–227.
- [Hestenes and Stiefel, 1952] Hestenes, M. R. and Stiefel, E. (1952). Methods of Conjugate Gradients for Solving Linear Systems. *Journal of Research of the National Bureau of Standards*, 49(6) :409–436.
- [Hill et al., 2001] Hill, D., Batchelor, P., Holden, M., and Hawkes, D. (2001). Medical image registration. *Physics in Medicine and Biology*, 46(3) :1–45.
- [Hill et al., 1994] Hill, D. L., Studholme, C., and Hawkes, D. J. (1994). Voxel similarity measures for automated image registration. In R. A. Robb, editor, *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 2359 of *Presented at the Society of Photo-Optical Instrumentation Engineers (SPIE) Conference*, pages 205–216.
- [Horn and Schunck, 1981] Horn, B. K. P. and Schunck, B. G. (1981). Determining optical flow. *Artificial Intelligence*, 17 :185–203.
- [Horsfield, 1999] Horsfield, M. A. (1999). Mapping eddy current induced fields for the correction of diffusion-weighted echo planar images. *Magnetic Resonance Imaging*, 17(9) :1335 – 1345.
- [Hsu and Ostuni, 1999] Hsu, L. Y. Loew, M. and Ostuni, J. (1999). Automated registration of brain images using edge and surface features. *IEEE Engineering in Medicine and Biology Magazine*, 18 :40–47.
- [Hutton and Braun, 2003] Hutton, B. F. and Braun, M. (2003). Software for image registration : Algorithms, accuracy, efficacy. *Seminars in Nuclear Medicine*, 33 :180–192.
- [Johnson and Christensen, 2002] Johnson, H. and Christensen, G. (2002). Consistent landmark and intensity-based image registration. *IEEE Transactions on Medical Imaging*, 21(5) :450–61.
- [Kiefer and Wolfowitz, 1952] Kiefer, J. and Wolfowitz, J. (1952). Stochastic estimation of the maximum of a regression function. *The Annals of Mathematical Statistics*, 23(3) :462–466.
- [Kim et al., 1996] Kim, B., Boes, J., Frey, K., and Meyer, C. (1996). Mutual information for automated multimodal image warping. In Höhne, K. and Kikinis, R., editors, *Visualization in Biomedical Computing*, volume 1131 of *Lecture Notes in Computer Science*, pages 349–354. Springer Berlin / Heidelberg.
- [Klein et al., 2009] Klein, A., Andersson, J., Ardekani, B., Ashburner, J., Avants, B., Chiang, M.-C., Christensen, G., Collins, D. L., Hellier, P., Song, J. H., Jenkinson, M., Lepage, C., Rueckert, D., Thompson, P., Vercauteren, T., Woods, R., Mann, J., and Parsey, R. (2009). Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. *NeuroImage*, 46(3) :786–802.
- [Klein et al., 2009] Klein, S., Pluim, J., Staring, M., and Viergever, M. (2009). Adaptive stochastic gradient descent optimisation for image registration. *International Journal of Computer Vision*, 81(3) :227–239.
- [Klein et al., 2007] Klein, S., Staring, M., and Pluim, J. P. W. (2007). Evaluation of optimization methods for nonrigid medical image registration using mutual information and b-splines. *IEEE Transactions on Image Processing*, 16(12) :2879–2890.
- [Kushner and Clark, 1978] Kushner, H. J. and Clark, D. S. (1978). *Stochastic approximation methods for constrained and unconstrained systems*. Springer-Verlag.
- [Kybic, 2001] Kybic, J. (2001). *Elastic Image Registration using Parametric Deformation Models*. PhD thesis, Ecole polytechnique fédérale de Lausanne, Suisse.

- [Lagarias et al., 1998] Lagarias, J. C., Reeds, J. A., Wright, M. H., and Wright, P. E. (1998). Convergence properties of the Nelder-Mead Simplex method in low dimensions. *SIAM Journal of Optimization*, 9 :112–147.
- [Lau et al., 2001] Lau, Y. H., Braun, M., and Hutton, B. F. (2001). Non-rigid image registration using a median-filtered coarse-to-fine displacement field and a symmetric correlation ratio. *Physics in Medicine and Biology*, 46(4) :1297.
- [Lester and Arridge, 1999] Lester, H. and Arridge, S. (1999). A survey of hierarchical non-linear medical image registration. *Pattern Recognition*, 32(1) :129–149.
- [Liu and Nocedal, 1989] Liu, D. C. and Nocedal, J. (1989). On the limited memory BFGS method for large scale optimization. *Mathematical Programming*, 45(3) :503–528.
- [Loeckx, 2006] Loeckx, D. (2006). *Automated nonrigid intra-patient image registration using B-splines*. PhD thesis, K.U Leuven, Belgique.
- [Maes et al., 1999] Maes, F., Vandermeulen, D., and Suetens, P. (1999). Comparative evaluation of multiresolution optimization strategies for multimodality image registration by maximization of mutual information. *Medical Image Analysis*, 3(4) :373 – 386.
- [Maintz and Viergever, 1998] Maintz, J. and Viergever, M. (1998). A survey of medical image registration. *Medical Image Analysis*, 2(1) :1–36.
- [Makela et al., 2002] Makela, T., Member, S., Clarysse, P., Sipilä, O., Pauna, N., Pham, Q. C., Katila, T., Magnin, I. E., and Axis, L. L. (2002). A review of cardiac image registration methods. *IEEE Transactions on Medical Imaging*, 21(9) :1011–1021.
- [Mangin, 1995] Mangin, J. (1995). *Mise en correspondance d’images médicales 3D multi-modalités multi-individus pour la corrélation anatomo-fonctionnelle cérébrale*. PhD thesis, Ecole Nationale Supérieure des Télécommunications, Paris, France.
- [McGillem and Svedlow, 1976] McGillem, C. and Svedlow, M. (1976). Image registration error variance as a measure of overlay quality. *IEEE Transactions on Geoscience Electronics*, 14(1) :44–49.
- [Mckinnon, 1999] Mckinnon, K. I. M. (1999). Convergence of the Nelder-Mead simplex method to a nonstationary point. *SIAM Journal on Optimization*, 9 :148–158.
- [McLaughlin et al., 2002] McLaughlin, R., Hipwell, J., Hawkes, D., Noble, J., Byrne, J., and Cox, T. (2002). A comparison of 2D-3D intensity-based registration and feature-based registration for neurointerventions. In Dohi, T. and Kikinis, R., editors, *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2002*, volume 2489 of *Lecture Notes in Computer Science*, pages 517–524. Springer Berlin / Heidelberg.
- [McLeish et al., 2002] McLeish, K., Hill, D., Atkinson, D., Blackall, J., and Razavi, R. (2002). A study of the motion and deformation of the heart due to respiration. *IEEE Transactions on Medical Imaging*, 21(9) :1142–1150.
- [Mellor and Brady, 2005] Mellor, M. and Brady, M. (2005). Phase mutual information as a similarity measure for registration. *Medical Image Analysis*, 9(4) :330 – 343. Functional Imaging and Modeling of the Heart - FIMH03, Functional Imaging and Modeling of the Heart.
- [Meyer et al., 1997] Meyer, C. R., Boes, J. L., Kim, B., Bland, P. H., Zasadny, K. R., Kison, P. V., Koral, K., Frey, K. A., and Wahl, R. L. (1997). Demonstration of accuracy and clinical versatility of mutual information for automatic multimodality image fusion using affine and thin-plate spline warped geometric deformations. *Medical Image Analysis*, 1(3) :195 – 206.
- [Modersitzki, 2004] Modersitzki, J. (2004). *Numerical Methods for Image Registration*. Oxford University Press.

- [Mora Cofre, 2008] Mora Cofre, M. A. (2008). *Ensembles de niveaux robustes au speckle et recalage B-spline : application à la segmentation et l'analyse du mouvement cardiaque par des images ultrasons*. PhD thesis, Institut de Recherche en Informatique de Toulouse, France.
- [Moré and Thuente, 1994] Moré, J. J. and Thuente, D. J. (1994). Line search algorithms with guaranteed sufficient decrease. *ACM Transactions on Mathematical Software*, 20(3) :286–307.
- [Nazareth and Tseng, 2002] Nazareth, L. and Tseng, P. (2002). Gilding the lily : A variant of the Nelder-Mead algorithm based on golden-section search. *Computational Optimization and Applications*, 22 :133–144.
- [Nelder, J. A. and Mead, R., 1965] Nelder, J. A. and Mead, R. (1965). A Simplex Method for Function Minimization. *The Computer Journal*, 7(4) :308–313.
- [Nielson, 1993] Nielson, G. M. (1993). Scattered data modeling. *IEEE Computer Graphics and Applications*, 13(1) :60–70.
- [Nikou et al., 1998a] Nikou, C., Heitz, F., Armspach, J., and Namer, I. (1998a). Single and Multimodal Subvoxel Registration of Dissimilar Medical Images using Robust Similarity Measures. In Hanson, K., editor, *Medical imaging 1998 : image processing*, volume 3338 of *SPIE Proceedings*, pages 167–178, Los Alamos National Lab., NM, USA.
- [Nikou et al., 1998b] Nikou, C., Heitz, F., and Armspach, J.-P. (1998b). Robust registration of dissimilar single and multimodal images. In Burkhardt, H. and Neumann, B., editors, *Computer Vision - ECCV98*, volume 1407 of *Lecture Notes in Computer Science*, pages 51–. Springer Berlin / Heidelberg.
- [Noblet, 2006] Noblet, V. (2006). *Recalage non rigide d'images cérébrales 3D avec contrainte de conservation de la topologie*. PhD thesis, Université Strasbourg-I, France.
- [Nocedal, 1980] Nocedal, J. (1980). Updating Quasi-Newton matrices with limited storage. *Mathematics of Computation*, 35(151) :773–782.
- [Nocedal and Wright, 2000] Nocedal, J. and Wright, S. J. (2000). *Numerical Optimization*. Springer.
- [Odobez and Bouthemy, 1995] Odobez, J. and Bouthemy, P. (1995). Robust multiresolution estimation of parametric motion models. *Journal of Visual Communication and Image Representation*, 6(4) :348–365.
- [Pelizzari et al., 1989] Pelizzari, C., Chen, G., Spelbring, D., Weichselbaum, R., and Chen, C. (1989). Accurate three-dimensional registration of ct, pet, and/or mr images of the brain. *Journal of Computer Assisted Tomography*, 13 :20–26.
- [Penney et al., 1998] Penney, G., Weese, J., Little, J., Desmedt, P., Hill, D., and Hawkes, D. (1998). A comparison of similarity measures for use in 2D-3D medical image registration. *IEEE Transactions on Medical Imaging*, 17(4) :586–595.
- [Petitjean, 2003] Petitjean, C. (2003). *Recalage non rigide d'images par approches variationnelles statistiques - Application à l'analyse et à la modélisation de la fonction myocardique en IRM*. PhD thesis, Université René Descartes - Paris V, France.
- [Plakhov and Cruz, 2004] Plakhov, A. and Cruz, P. (2004). A stochastic approximation algorithm with step-size adaptation. *Journal of Mathematical Sciences*, 120 :964–973.
- [Pluim et al., 2000] Pluim, J., Maintz, J., and Viergever, M. (2000). Image registration by maximization of combined mutual information and gradient information. *IEEE Transactions on Medical Imaging*, 19(8) :809–814.

- [Pluim et al., 2003] Pluim, J., Maintz, J., and Viergever, M. (2003). Mutual information based registration of medical images : a survey. *IEEE Transactions on Medical Imaging*, 22(8) :986–1004.
- [Potra and Shi, 1995] Potra, F. and Shi, Y. (1995). Efficient line search algorithm for unconstrained optimization. *Journal of Optimization Theory and Applications*, 85 :677–704.
- [Roche, 2001] Roche, A. (2001). *Recalage d'images médicales par inférence statistique*. PhD thesis, Université de Nice - Sophia Antipolis, France.
- [Roche et al., 1999] Roche, A., Malandain, G., and Ayache, N. (1999). Unifying Maximum Likelihood Approaches in Medical Image Registration. Technical Report 3741, INRIA.
- [Roche et al., 1998] Roche, A., Malandain, G., Pennec, X., and Ayache, N. (1998). The Correlation Ratio as a New Similarity Measure for Multimodal Image Registration. In W.M. Wells, A. C. and Delp, S., editors, *Medical Image Computing and Computer-Assisted Intervention : MICCAI 1998*, volume 1496, pages 1115–1124. lncs.
- [Rohde et al., 2003] Rohde, G. K., Member, S., Aldroubi, A., Dawant, B. M., and Member, S. (2003). The adaptive bases algorithm for intensity-based nonrigid image registration. *IEEE Transactions on Medical Imaging*, 22 :1470–1479.
- [Rohlfing et al., 2003] Rohlfing, T., Maurer, C. J., Bluemke, D., and MA, J. (2003). Volume-preserving nonrigid registration of MR breast images using free-form deformation with an incompressibility constraint. *IEEE Transactions on Medical Imaging*, 22(6) :730–41.
- [Rohr, 2000] Rohr, K. (2000). Elastic registration of multimodal medical images : A survey. *Kunstliche Intelligenz*, 3 :11–17.
- [Rohr et al., 1999] Rohr, K., Fornefett, M., and Stiehl, H. S. (1999). Approximating thin-plate splines for elastic registration : Integration of landmark errors and orientation attributes. In *Proceedings of Information Processing in Medical Imaging : IPMI 1999, volume 1613 of LNCS*, pages 252–265. Springer.
- [Rousseau, 2003] Rousseau, F. (2003). *Méthodes d'analyse d'images et de calibration pour l'échographie 3D en mode main-libre*. PhD thesis, Université de Rennes 1, IFSIC, IRISA, Rennes, France.
- [Rueckert et al., 1999] Rueckert, D., Sonoda, L., Hayes, C., Hill, D., Leach, M., and Hawkes, D. (1999). Nonrigid registration using free-form deformations : Application to breast MR images. *IEEE Transactions on Medical Imaging*, 18(8) :712–721.
- [Sarrut, 2000] Sarrut, D. (2000). *Recalage multimodal et plate-forme d'imagerie médicale à accès distant*. PhD thesis, Université Lumière, Lyon 2, France.
- [Schnabel et al., 2001] Schnabel, J., Rueckert, D., Quist, M., Blackall, J., Castellano-Smith, A., Hartkens, T., Penney, G., Hall, W., Liu, H., Truwit, C., Gerritsen, F., Hill, D., and Hawkes, D. (2001). A generic framework for non-rigid registration based on non-uniform multi-level free-form deformations. In Niessen, W. and Viergever, M. A., editors, *Medical Image Computing and Computer-Assisted Intervention MICCAI'2001*, volume 2208, pages 573–581, Utrecht (Netherlands).
- [Schnabel et al., 2003] Schnabel, J., Tanner, C., AD., C.-S., Degenhard, A., Leach, M., Hose, D., DL., H., and Hawkes, D. (2003). Validation of nonrigid image registration using finite-element methods : application to breast MR images. *IEEE Transactions on Medical Imaging*, 22(2) :238–47.
- [Sederberg and Parry, 1986] Sederberg, T. W. and Parry, S. R. (1986). Free-form deformation of solid geometric models. In *SIGGRAPH '86 : Proceedings of the 13th annual conference on Computer graphics and interactive techniques*, pages 151–160, New York, NY, USA. ACM Press.

- [Shams et al., 2010] Shams, R., Sadeghi, P., Kennedy, R., and Hartley, R. (2010). A survey of medical image registration on multicore and the GPU. *IEEE Signal Processing Magazine*, 27(2) :50–60.
- [Shanno, 1970] Shanno, D. F. (1970). Conditioning of quasi-newton methods for function minimization. *Mathematics of Computation*, 24 :647–656.
- [Shannon, 1948] Shannon, C. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27 :379–423/623–656.
- [Shen and Davatzikos, 2002] Shen, D. and Davatzikos, C. (2002). Hammer : Hierarchical attribute matching mechanism for elastic registration. *IEEE Transactions on Medical Imaging*, 21 :1421–1439.
- [Simon, 2005] Simon, A. (2005). *Extraction et caractérisation du mouvement cardiaque en imagerie scanner multibarrette*. PhD thesis, LTSI, Université Rennes 1, France.
- [Simon et al., 2005] Simon, A., Garreau, M., Boulmier, D., Coatrieux, J.-L., and Breton, H. L. (2005). A surface-volume matching process using a markov random field model for cardiac motion extraction in MSCT imaging. pages 457–466.
- [Simon et al., 2006] Simon, A., Garreau, M., Boulmier, D., Coatrieux, J.-L., and Breton, H. L. (2006). Estimation combiné de forme et de mouvement par champs markoviens : application À l'imagerie cardiaque scanner multibarrette. *Traitement du Signal*, 23(5-6) :473–487.
- [Sorzano and Unser, 2005] Sorzano, C. O. S. Thevenaz, P. and Unser, M. (2005). Elastic registration of biological images using vector-spline regularization. *IEEE transactions on biomedical engineering*, 52 :652–663.
- [Stiller and Konrad, 1999] Stiller, C. and Konrad, J. (1999). Estimating motion in image sequences. *IEEE Magazine on Signal Processing*, 16 :70–91.
- [Studholme et al., 1999] Studholme, C., Hill, D. L. G., and Hawkes, D. J. (1999). An overlap invariant entropy measure of 3D medical image alignment. *Pattern Recognition*, 32(1) :71–86.
- [Studholme and Hawkes, 1995] Studholme, C. Hill, D. L. G. and Hawkes, D. J. (1995). Multiresolution voxel similarity measures for MR-PET registration. In Bizais, Y. and Barillot, C., editors, *Information processing in medical imaging*, Computational imaging and vision, pages 287–298.
- [Svedlow et al., 1978] Svedlow, M., McGillem, C., and Anuta, P. (1978). Image registration : Similarity measure and processing method comparisons. 14(1) :141–149.
- [Svedlow and Anuta, 1976] Svedlow, M. M. C. D. and Anuta, P. E. (1976). Experimental examination of similarity measures and preprocessing methods used for image registration. In *Symposium on machine processing of remotely sensed data*.
- [Swain and Ballard, 1991] Swain, M. J. and Ballard, D. H. (1991). Color indexing. *International Journal of Computer Vision*, 7 :11–32.
- [Talairach and Tournoux, 1988] Talairach, J. and Tournoux, P. (1988). *Co-Planar Stereotaxic Atlas of the Human Brain*. Thieme.
- [Tanner et al., 2002] Tanner, C., Schnabel, J., Degenhard, A., Castellano-Smith, A., Hayes, C., Leach, M., Hose, D., Hill, D., and Hawkes, D. (2002). Validation of volume-preserving non-rigid registration : Application to contrast-enhanced MR-mammography. In Dohi, T. and Kikinis, R., editors, *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2002*, volume 2488 of *Lecture Notes in Computer Science*, pages 307–314. Springer Berlin / Heidelberg.

- [Tavard et al., 2010] Tavard, F., Simon, A., Donal, E., Hernandez, A., and Garreau, M. (2010). Fusion of electro-anatomical mapping and speckle tracking echography for the characterization of local electro-mechanical delays in crt optimization.
- [Tavard et al., 2009] Tavard, F., Simon, A., Leclercq, C., Pavin, D., Hernandez, A., and Garreau, M. (2009). Data fusion of left ventricle electro-anatomical mapping and multislice computerized tomography. pages 1745–1748.
- [Tseng, 1999] Tseng, P. (1999). Fortified-descent simplicial search method : A general approach. *SIAM Journal on Optimization*, 10 :269–288.
- [Veltkamp and Hagedoorn, 1999] Veltkamp, R. C. and Hagedoorn, M. (1999). State-of-the-art in shape matching. Technical report, Principles of Visual Information Retrieval.
- [Venot et al., 1984] Venot, A., Lebruchec, J., and Roucayrol, J. (1984). A new class of similarity measures for robust image registration. *Computer Vision, Graphics, and Image Processing*, 28(2) :176 – 184.
- [Viola, 1995] Viola, P. (1995). *Alignment by Maximization of Mutual Information*. PhD thesis, Massachusetts Institute of Technology Artificial Intelligence Laboratory, Boston, USA.
- [Walters et al., 1991] Walters, F., Parker, L., Morgan, S., and Deming, S. (1991). *Sequential Simplex Optimization*. CRC Press.
- [West et al., 1997] West, J., Fitzpatrick, J. M., Wang, M. Y., Dawant, B. M., Maurer, C. R., Kessler, R. M., Maciunas, R. J., Barillot, C., Lemoine, D., Collignon, A., Maes, F., Sumanaweera, T. S., Harkness, B., Hemler, P. F., Hill, D. L. G., Hawkes, D. J., Studholme, C., Maintz, J. B. A., Viergever, M. A., Mal, G., Pennec, X., Noz, M. E., Maguire, G. Q., Pollock, M., Pelizzari, C. A., Robb, R. A., Hanson, D., and Woods, R. P. (1997). Comparison and Evaluation of Retrospective Intermodality Image Registration Techniques. *Journal of Computer Assisted Tomography*, 21(4) :554–566.
- [Woods et al., 1993] Woods, R., Mazziota, J., and Cherry, S. (1993). MRI-PET Registration with Automated Algorithm. *Journal of Computer Assisted Tomography*, 17(4) :536–546.
- [Xue and Davatzikos, 2004] Xue, Z. Shen, D. and Davatzikos, C. (2004). Determining correspondence in 3-D MR brain images using attribute vectors as morphological signatures of voxels. *IEEE Transactions on Medical Imaging*, 23 :1276–1291.
- [Yuille et al., 1992] Yuille, A. L., Hallinan, P. W., and Cohen, D. S. (1992). Feature extraction from faces using deformable templates. *International Journal of Computer Vision*, 8 :99–111.
- [Zitova and Flusser, 2003] Zitova, B. and Flusser, J. (2003). Image registration methods : a survey. *Image and Vision Computing*, 21 :977–1000.

Chapitre 2

Information Mutuelle en recalage d'images médicales

L'Information Mutuelle (IM) est née du fruit de travail de Claude Elwood Shannon en 1949 [Shannon and Weaver, 1949], peu après avoir introduit l'entropie différentielle qui porte maintenant son nom [Shannon, 1948]. En recalage d'images, l'IM fut utilisée pour la première fois et la même année par deux équipes de recherche [Collignon et al., 1995a, Viola, 1995]. Depuis 1995, la liste des publications qui traitent de cette mesure en recalage ne cesse de s'allonger, et elle est devenue La mesure de référence en recalage d'images médicales multi-modales [Meyer et al., 1997, Pluim et al., 2003]. Dans ce chapitre, nous allons présenter l'IM et ses différentes versions proposées dans la littérature. Celles-ci diffèrent par leur formulation aussi bien que par la méthode d'estimation utilisée, et tendent à minimiser les défauts inhérents à l'IM classique. Nous commençons par faire quelques rappels de théorie de l'information nécessaires à la compréhension de cette mesure. Puis, nous présentons les méthodes fondatrices d'estimation qui ont été développées. En nous basant sur les limitations avérées des estimateurs classiques l'IM, nous ferons ensuite un tour d'horizon des nouvelles méthodes qui ont été mises en oeuvre ces dernières années, et qui tendent à se prémunir de ses limitations, notamment en proposant de nouveaux estimateurs de l'IM. Nous essaierons de classer ces nouvelles méthodes en fonction de leur efficacité. Cependant, cette tâche est rendue difficile par le fait qu'il n'existe que quelques rares études comparatives globales de ces méthodes [Pluim et al., 2004, Škerl et al., 2007, Cahill, 2010].

2.1 Entropie de Shannon

La première mesure de l'information a été donnée par Hartley en 1928 [Hartley, 1928], un électronicien américain qui contribua aux fondations de la théorie de l'information. Sa mesure de l'entropie est donnée par :

$$H(x) = \log s \quad (2.1)$$

où s le nombre de valeurs possibles pour chaque symbole du message x transmis. Cette mesure d'entropie est basée sur l'hypothèse que tous les événements sont équiprobables, hypothèse régulièrement mise en défaut dans les applications de télécommunications ou de traitement de signal. Pour pallier ce problème, Shannon modifia la mesure de Hartley en pondérant chaque événement par son nombre d'occurrences [Shannon, 1948]. Shannon donna ainsi naissance à l'entropie qui porte maintenant son nom, une fonction mathématique qui correspond à la quantité d'information contenue ou délivrée par une source d'information. Originellement développée pour formaliser la nature statistique de l'information perdue dans les signaux des

lignes téléphoniques, cette mesure très informative a ensuite été reprise notamment en informatique, où elle est toujours utilisée de nos jours pour quantifier, par exemple, le nombre de bits sur lesquels on peut coder un fichier sans perte de données. L'entropie de Shannon, ou entropie différentielle, d'une VA x est ainsi donnée par :

$$H(x) = - \int_{\mathbb{R}} p_x(u) \log(p_x(u)) du \quad (2.2)$$

où p_x représente la densité de probabilité de la VA x . L'entropie de Shannon d'une image peut être estimée à partir des intensités de l'image et plus précisément de la distribution de ses niveaux de gris. Comme nous allons le voir, les densités de probabilités sont estimées à partir des histogrammes des images considérées. Si nous voulons expliciter la signification de l'entropie en traitement d'images, nous dirons qu'une image constituée de seulement quelques niveaux de gris différents aura une entropie faible, c'est à dire qu'elle « contiendra peu d'information », tandis qu'une image contenant beaucoup d'intensités différentes, avec des occurrences équivalentes, aura une entropie élevée, et contiendra donc plus d'information. L'entropie de Shannon peut donc être considérée comme une mesure de dispersion d'une densité de probabilité, une distribution avec un pic prédominant entraînant une entropie faible, alors qu'une distribution plus étalée donne une entropie plus importante. Dans la suite, quand nous ne le préciserons pas, nous ferons toujours référence à l'entropie de Shannon.

2.2 Histogramme conjoint et entropie conjointe

Tout en conservant le formalisme présenté dans le chapitre 1, nous allons maintenant introduire le concept d'histogramme conjoint, qui est au centre de l'estimation de l'IM. L'histogramme conjoint $h_{\mathbf{x}}$ du vecteur aléatoire \mathbf{x} est défini de $D^{(1)} \times D^{(2)}$ dans \mathbb{R}^+ . Il consiste à estimer la distribution de probabilité conjointe du vecteur aléatoire $\mathbf{x} = (x^{(1)}, x^{(2)})^T$ en comptabilisant les occurrences des valeurs prises par le vecteur sur l'ensemble de ses réalisations. Le tirage aléatoire de $x^{(1)}$ (resp. $x^{(2)}$) représentant l'intensité du pixel ν de l'image de référence (resp. flottante) sera alors notée $x^{(1)}(\nu)$ (resp. $x^{(2)}(\nu)$). En notant $\mathbf{x}(\nu) = (x^{(1)}(\nu), x^{(2)}(\nu))^T$ l'ensemble des réalisations du vecteur aléatoire \mathbf{x} , on a donc :

$$\forall \mathbf{u} \in D^{(1)} \times D^{(2)}, h_{\mathbf{x}}(\nu) = \text{Card}\{\mathbf{x}(\nu) = \mathbf{u}\} \quad (2.3)$$

où $\mathbf{u} = (u^{(1)}, u^{(2)})^T$ correspond aux couples d'intensités possibles. Une représentation graphique de l'histogramme conjoint associé à un couple d'images est donnée sur la figure 2.1. Nous donnons également l'exemple de 2 histogrammes conjoints avant et après recalage sur la figure 2.2. Comme nous l'avons dit plus haut, et ce sous certaines hypothèses, la **densité de probabilité conjointe** du vecteur aléatoire \mathbf{x} peut être estimée à l'aide de l'histogramme conjoint $h_{\mathbf{x}}$ de \mathbf{x} :

$$p_{\mathbf{x}} \approx \frac{h_{\mathbf{x}}}{\sum_{\mathbf{u}} h_{\mathbf{x}}} \quad (2.4)$$

Les **densités marginales**, quant à elles, peuvent être estimées en sommant les lignes ou les colonnes de l'histogramme conjoint normalisé. Ainsi, nous pouvons écrire :

$$p_{x^{(1)}}(u^{(1)}) \approx \sum_{u^{(2)}} h_{\mathbf{x}}(u^{(1)}, u^{(2)}) \quad (2.5)$$

$$p_{x^{(2)}}(u^{(2)}) \approx \sum_{u^{(1)}} h_{\mathbf{x}}(u^{(1)}, u^{(2)}) \quad (2.6)$$

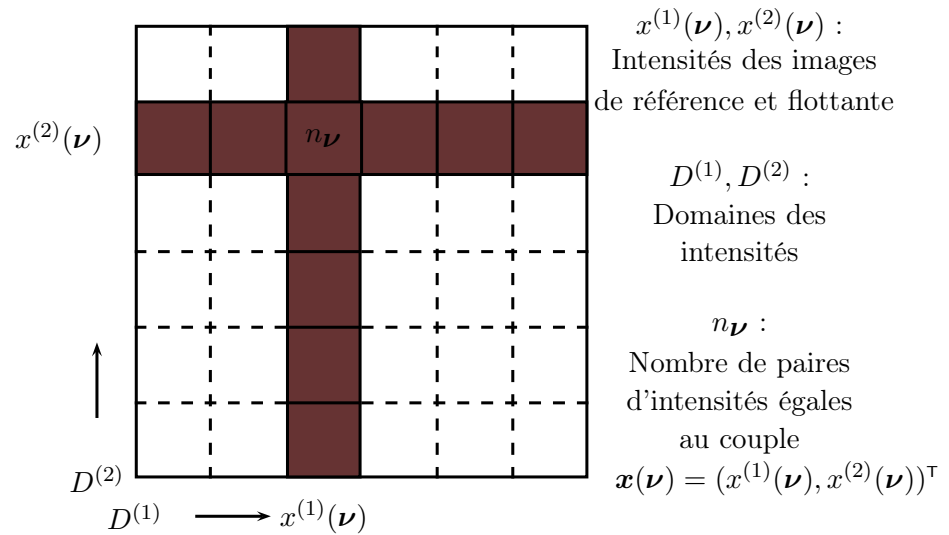


FIGURE 2.1 – Représentation graphique de l'histogramme conjoint d'une image de référence (1) et d'une image flottante (2).

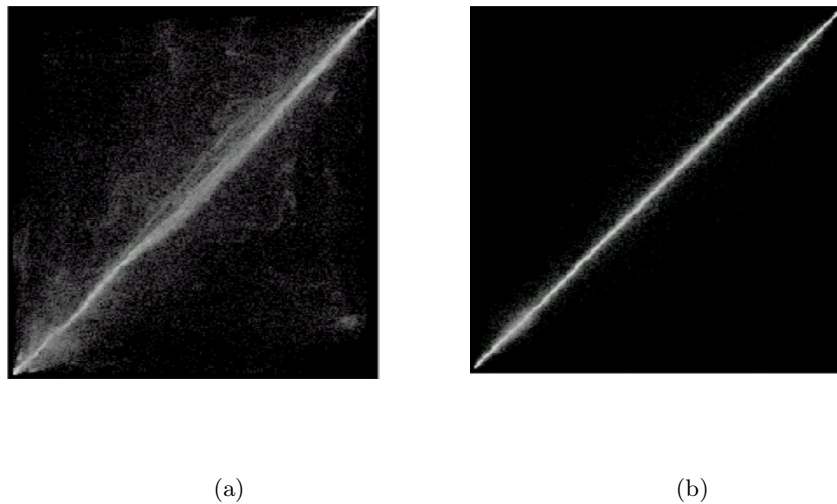


FIGURE 2.2 – Exemple d'histogrammes conjoints obtenus pour (a) des images non recalées et (b) des images recalées.

La grande avancée en recalage d'images que l'on doit à [Collignon et al., 1995b], [Studholme and Hawkes, 1995], est d'avoir utilisé l'**entropie conjointe de Shannon**. Sa formule est :

$$H(\mathbf{x}) = - \oint_{\mathbb{R}^2} p_{\mathbf{x}}(\mathbf{u}) \log(p_{\mathbf{x}}(\mathbf{u})) d\mathbf{u} \quad (2.7)$$

Elle est donc fonction de $p_{\mathbf{x}}$, la densité de probabilité conjointe des 2 images, qui peut elle-même être estimée à partir de l'histogramme conjoint (voir (2.4)). Recaler deux images revient alors à trouver la transformation qui minimise leur entropie conjointe.

2.3 Information mutuelle

A partir du moment où l'histogramme conjoint fut introduit pour le recalage d'images médicales multimodales [Collignon et al., 1995a, Viola, 1995], l'IM ne tarda pas à faire son apparition. Différentes formulations de l'IM ont été données dans la littérature. Trois d'entre-elles sont souvent utilisées, et bien qu'identiques mathématiquement, elles permettent d'appréhender le problème du recalage de façon différente. Toutefois, dans les 3 cas, le problème du recalage revient à maximiser l'IM. La première de ces formulations est :

$$\begin{aligned} IM(\mathbf{x}) &= H(x^{(2)}) - H(x^{(2)}|x^{(1)}) \\ &= H(x^{(1)}) - H(x^{(1)}|x^{(2)}) \end{aligned} \quad (2.8)$$

où $H(x^{(2)})$ est l'entropie de Shannon de la VA $x^{(2)}$ et $H(x^{(2)}|x^{(1)})$ est l'entropie conditionnelle calculée à partir de la densité de probabilité conditionnelle $p_{x^{(2)}|x^{(1)}}$. Avec cette formulation, l'IM représente la quantité par laquelle l'incertitude par rapport à $x^{(2)}$ décroît lorsque $x^{(1)}$ est donnée, autrement dit la quantité d'information de $x^{(1)}$ contenue dans $x^{(2)}$. Comme les 2 VAs peuvent être interchangées, $IM(\mathbf{x})$ représente également la quantité d'information de $x^{(2)}$ contenue dans $x^{(1)}$. On a donc bien une information mutuelle entre les images à recaler, et les images doivent être alignées de façon à ce que la quantité d'information commune soit maximale.

Une deuxième interprétation possible est donnée par la formulation la plus classique de l'IM dans la littérature, qui est :

$$IM(\mathbf{x}) = H(x^{(1)}) + H(x^{(2)}) - H(\mathbf{x}) \quad (2.9)$$

Ici, l'entropie conjointe apparaît explicitement dans l'expression de l'IM : Maximiser l'IM revient alors à minimiser l'entropie conjointe plus un terme de pénalité donné par l'entropie marginale de $x^{(2)}$ (car dans le contexte du recalage, l'entropie marginale de $x^{(1)}$, associée à l'image de référence, reste constante en fonction de la transformation considérée). L'avantage de l'IM par rapport à l'entropie conjointe est ainsi l'ajout de ce terme correspondant à l'entropie marginale de l'image à recaler. En effet, l'entropie conjointe a un inconvénient majeur, c'est qu'elle est très sensible au recouvrement partiel. Prenons un exemple pour illustrer ce problème. Si le recalage diverge et que la transformation aberrante qui en découle fait « sortir l'image de son cadre », l'image flottante résultante sera totalement noire (ou du moins elle prendra la valeur par défaut des pixels), ce qui donnera un histogramme conjoint avec un seul pic, celui de ce niveau de gris noir. Alors, au sens de l'entropie conjointe, l'algorithme aura trouvé un optimum. L'entropie marginale de l'image à recaler (ou flottante) parvient à

s'affranchir de ce problème en contrecarrant d'une certaine façon l'entropie conjointe : elle pénalise les transformations qui font décroître la quantité d'information dans l'image flottante prise séparément. Ainsi, l'IM est un peu plus robuste que l'entropie conjointe face au problème de recouvrement partiel.

La dernière formulation de l'IM utilise explicitement la **divergence de Kullback-Leibler** (KL), qui est définie par :

$$KL(p_{\mathbf{x}}, p_{\mathbf{y}}) = \oint p_{\mathbf{x}}(\mathbf{u}) \log \left(\frac{p_{\mathbf{x}}(\mathbf{u})}{p_{\mathbf{y}}(\mathbf{u})} \right) d\mathbf{u} \quad (2.10)$$

Cette dernière mesure l'écart entre deux densités de probabilité. On a alors :

$$IM(\mathbf{x}) = KL(p_{\mathbf{x}}, p_{x^{(1)}}p_{x^{(2)}}) \quad (2.11)$$

Les deux VAs $x^{(1)}$ et $x^{(2)}$ étant indépendantes si et seulement si $p_{\mathbf{x}} = p_{x^{(1)}}p_{x^{(2)}}$. Cette formulation montre clairement que l'IM peut être utilisée comme mesure de dépendance entre 2 images. La dépendance entre les niveaux de gris des images est maximale lorsqu'elles sont correctement alignées.

A travers les 3 définitions précédentes, on peut mieux appréhender le rôle de l'IM comme mesure de similarité entre images pour le recalage multimodal. On comprend notamment mieux l'hypothèse de dépendance statistique : aucune hypothèse, dans ces définitions, n'est en effet faite sur la nature même des images.

Quelle que soit la définition appliquée, la mesure de l'IM nécessite en pratique l'estimation des probabilités marginales et conjointe, et donc classiquement la construction d'histogrammes marginaux et conjoint. Beaucoup de travaux ont porté spécifiquement sur le problème de la construction d'histogrammes. Dans ce qui suit, nous allons donc présenter les différentes méthodes couramment utilisées pour estimer une densité de probabilité, car c'est un point essentiel d'une méthode de recalage basée sur l'IM : de la qualité d'estimation des densités de probabilité dépend directement la qualité de l'estimation de l'IM, et donc celle du recalage.

2.4 Méthodes d'estimation de référence

On peut donner 3 grandes familles d'estimateurs de l'IM [Moddemeijer, 1999] : **les estimateurs à histogramme**, **les estimateurs à noyaux** (ces 2 premiers étant non paramétriques), et enfin **les estimateurs paramétriques**. Seules les 2 premières familles nous intéressent en pratique, car les méthodes paramétriques sont trop coûteuses à mettre en oeuvre dans le cadre du recalage d'images médicales. Les estimateurs à histogramme sont les plus utilisés en pratique, car leur conception est plus simple. Mais ils souffrent de problèmes tels un biais qui peut être important et que les estimateurs à noyaux tendent à minimiser. Notons qu'il n'existe pas de consensus quant à la méthode d'estimation, car tous les aspects de l'implémentation interagissent dans l'algorithme de recalage (estimation des densités de probabilités, interpolation, optimisation, etc.). Dans la littérature du recalage, trois méthodes de base ont été principalement utilisées pour estimer une densité de probabilité et donc l'IM : l'estimateur à histogramme, l'estimateur à noyau gaussien et l'estimateur à noyau B-Spline. Dans ce qui suit, nous présentons et donnons les avantages et inconvénients de ces trois méthodes.

2.4.1 Estimateur à histogramme

Grâce à leur simplicité d'utilisation et à leurs propriétés, les estimateurs à histogramme sont les plus utilisés [Pluim et al., 2003]. Leurs propriétés en terme de biais, variance ou robustesse,

sont en outre relativement bien connues [Moddemeijer, 1999]. C'est d'ailleurs cet estimateur qui a été choisi dans un des deux articles originels utilisant pour la première fois l'IM comme mesure de similarité en recalage d'images médicales [Collignon et al., 1995a]. L'histogramme permet de représenter la répartition d'une ou plusieurs variables. Sa construction nécessite un choix important : celui du nombre de classes utilisées sur la plage de valeurs codant les niveaux de gris de nos images (appelée dynamique des niveaux de gris). C'est en lui-même un problème assez délicat. En effet, un nombre de classes non-optimal amène à une estimation biaisée : si le nombre de classes est insuffisant, l'histogramme sera trop lissé, entraînant une perte d'information (c'est regrettable quand on a choisi une méthode iconique justement pour avoir accès à toute l'information disponible dans l'image), tandis qu'un nombre de classes trop important fera ressortir les détails des images. Dans [Collignon et al., 1995a], la définition de l'IM choisie est celle liée à la divergence de Kullback-Leibler (2.11). L'estimation de la densité de probabilité conjointe se fait simplement par normalisation de l'histogramme conjoint par :

$$p_x \approx \frac{h_{\tilde{x}}}{\text{Card}(\mathcal{V})} \quad (2.12)$$

avec $\tilde{x}^{(i)}(\nu) = \frac{x^{(i)}(\nu) - x_{\min}^{(i)}}{\Delta b}$ les versions normalisées des niveaux de gris de l'image : $x_{\min}^{(i)} = \min_{\nu} (x^{(i)}(\nu))$ représente l'intensité minimale de l'image et Δb correspond à la plage d'intensités de chaque classe de l'histogramme. Enfin, \mathcal{V} représente les pixels/voxels pris en compte dans l'estimation. Dans [Collignon et al., 1995a], peu de détails concernant le choix du nombre de classes de l'histogramme sont donnés. Pour en apprendre davantage, il faut s'intéresser à la méthode d'estimation à partir d'histogrammes proposée dans l'article [Studholme et al., 1997], qui est l'une des plus usitées. L'évaluation de la densité de probabilité conjointe est notamment détaillée : la plage d'intensités de l'image considérée est calculée. Puis cette plage est divisée par deux jusqu'à ce qu'elle soit inférieure à 128 (et supérieure à 64). Cette plage normalisée permet ensuite de fixer le nombre de classes de l'histogramme. Cette manière de faire est valable pour la plus haute résolution de l'image. Quant l'image est sous-échantillonnée, le nombre de classes est divisé proportionnellement au nombre de pixels. L'estimation des densités de probabilité marginales est obtenue en effectuant une sommation par rapport aux lignes ou aux colonnes de l'histogramme conjoint. Le constat fait par Studholme et al. est que prendre 64 ou 256 classes à haute résolution ne modifie pas fondamentalement les performances. Par contre, un grand nombre de classes augmente significativement le temps de calcul puisqu'il implique une augmentation de la taille de l'histogramme conjoint dont la construction représente une grande partie du temps de calcul lors d'une procédure de recalage. Cet estimateur est utilisé dans un grand nombre de méthodes [Maes et al., 1997, Studholme et al., 1999, Rueckert et al., 1999, Slomka et al., 2003, Rohlfing et al., 2004], etc.

2.4.2 Estimateur à noyau

On compte plusieurs estimateurs à noyaux, autrement appelés estimateurs par fenêtrage de Parzen [Rosenblatt, 1956, Parzen, 1962], qui sont largement répandus dans la littérature du recalage. Parmi ceux-ci, l'estimateur à noyau gaussien est le premier à être apparu [Viola, 1995]. Quel que soit le noyau utilisé, le principe reste le même : au lieu de prendre un pixel et de lui attribuer un poids de 1 dans l'histogramme comme c'est le cas pour les estimateurs précédemment cités, l'influence du pixel s'étend par le biais du noyau à ses voisins. Une illustration de cette méthode, comparée à l'estimation par histogramme, est donnée figure 2.3. Cela permet d'obtenir des histogrammes lissés, et par conséquent moins sensibles aux nombres de classes préalablement choisis. **L'estimateur de Parzen** d'une densité de probabilité p_x peut s'écrire :

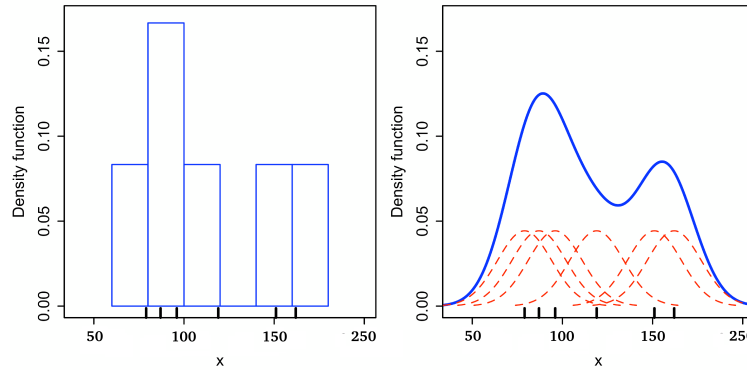


FIGURE 2.3 – Comparaison entre l'estimation par histogramme (à gauche) et celle par fenêtrage de Parzen (à droite). Ici, le noyau employé est gaussien. Les lignes verticales en noir (au niveau de l'abscisse) représentent les données, et les courbes en pointillé rouge représentent les différents noyaux utilisés lors de l'estimation. On constate qu'avec les mêmes données de départ, la densité de probabilité estimée (en bleu) est très différente suivant l'estimateur utilisé.

$$p_x(u) \approx \frac{1}{\text{Card}(\mathcal{V})\Delta b} \sum_{\nu \in \mathcal{V}} K\left(\frac{u - x(\nu)}{\Delta b}\right) \quad (2.13)$$

avec K le noyau utilisé, et Δb le paramètre de lissage, la largeur de la fenêtre, choisi par l'utilisateur. Comme pour l'estimateur par histogramme, une petite valeur de Δb a tendance à ne pas lisser suffisamment les données, ce qui conduit à l'apparition de détails inutiles, tandis qu'une valeur de Δb trop grande va au contraire éliminer tous les détails. Cette méthode d'estimation tient sa popularité de trois points majeurs :

- l'expression théorique (2.13) est très simple, puisque l'estimation empirique de p_x se ramène à une somme de VA.
- l'estimation de p_x converge vers la densité de probabilité, notamment au sens $L1$ quand $1/\Delta b$ et $\text{Card}(\mathcal{V})\Delta b$ tendent vers l'infini. Ceci est très important puisque la qualité de l'estimation dépendra directement du nombre de réalisations de la VA pris en compte. Dans notre cas, elle dépendra donc du nombre de pixels de l'image utilisés pour l'estimation de cette densité de probabilité. Nous reviendrons très largement sur ce point dans la suite, et notamment dans le chapitre 4, où une étude de la précision des estimateurs de l'IM en fonction du nombre d'échantillons est menée.
- cette méthode d'estimation est flexible puisqu'elle laisse à l'utilisateur un large choix de noyaux K . Nous détaillons les deux noyaux les plus usités en recalage dans ce qui suit.

2.4.3 Estimateur à noyau gaussien

En théorie, le noyau sinus cardinal est le meilleur puisqu'il a un support de largeur infinie. En pratique cependant, il serait trop coûteux à mettre en oeuvre. Viola et Wells [Viola and Wells, 1997] se sont donc tournés vers le noyau gaussien. Dans cet article originel de l'utilisation de l'IM en recalage, l'entropie est donnée comme l'opposé de l'espérance du logarithme de la densité de probabilité (écriture équivalente à celle de l'équation (2.7)) :

$$H(x) = -\mathbb{E}[\log(p_x(x))] \quad (2.14)$$

L'estimation des densités de probabilité se fait par fenêtrage de Parzen avec un noyau gaussien. La densité de probabilité marginale est estimée par :

$$p_x(u) \approx \frac{1}{Card(\mathcal{V}_a)} \sum_{\boldsymbol{\nu} \in \mathcal{V}_a} \phi_x^N(u - x(\boldsymbol{\nu})) \quad (2.15)$$

où \mathcal{V}_a représente l'ensemble a des pixels/voxels pris en compte dans l'estimation. Elle consiste en une superposition de distributions gaussiennes ϕ_x^N centrées sur $Card(\mathcal{V}_a)$ réalisations $x(\boldsymbol{\nu})$ (les niveaux de gris) de la VA x (associée à l'image). L'estimée de p_x correspond ainsi à la somme des contributions de chaque pixel $\boldsymbol{\nu} \in \mathcal{V}_a$, contributions qui dépendent de la distance entre u et $x(\boldsymbol{\nu})$. Comme il est précisé dans [Viola and Wells, 1997], l'utilisation de densités gaussiennes simplifie les calculs, mais n'est pas nécessaire. En fait, toute fonction différentiable peut être utilisée : une densité de Cauchy est également un bon choix de noyau.

L'approximation de l'espérance est quant à elle obtenue par une moyenne empirique :

$$\mathbb{E}[f(x)] \approx \frac{1}{Card(\mathcal{V}_b)} \sum_{\boldsymbol{\nu} \in \mathcal{V}_b} f(x(\boldsymbol{\nu})) \quad (2.16)$$

Il faut noter que les ensembles \mathcal{V}_a et \mathcal{V}_b des pixels utilisés pour le calcul de (2.15) et (2.16) sont différents. Ainsi, l'approximation de l'entropie (2.14) peut s'exprimer, à partir de (2.15) et (2.16), par :

$$H(x) \approx \frac{-1}{Card(\mathcal{V}_b)} \sum_{\boldsymbol{\nu}_j \in \mathcal{V}_b} \log \left(\frac{1}{Card(\mathcal{V}_a)} \sum_{\boldsymbol{\nu}_i \in \mathcal{V}_a} \phi_x^N(x(\boldsymbol{\nu}_j) - x(\boldsymbol{\nu}_i)) \right) \quad (2.17)$$

Une approximation de la dérivée de l'IM est utilisée pour mener la procédure d'optimisation. Pour obtenir cette dérivée, il faut tout d'abord donner la dérivée de l'estimée de l'entropie par rapport aux paramètres $\boldsymbol{\mu}$ de la transformation. Celle-ci s'écrit :

$$\frac{d}{d\boldsymbol{\mu}} H(x) \approx \frac{1}{Card(\mathcal{V}_b)} \sum_{\boldsymbol{\nu}_j \in \mathcal{V}_b} \sum_{\boldsymbol{\nu}_i \in \mathcal{V}_a} W_x(x(\boldsymbol{\nu}_i), x(\boldsymbol{\nu}_j)) (x(\boldsymbol{\nu}_i) - x(\boldsymbol{\nu}_j)) \sigma_x^{-1} \frac{d}{d\boldsymbol{\mu}} (x(\boldsymbol{\nu}_i) - x(\boldsymbol{\nu}_j)) \quad (2.18)$$

où σ_x^{-1} est l'inverse de la variance de x et :

$$W_x(x(\boldsymbol{\nu}_i), x(\boldsymbol{\nu}_j)) = \frac{\phi_x^N(x(\boldsymbol{\nu}_i) - x(\boldsymbol{\nu}_j))}{\sum_{\boldsymbol{\nu}_k \in \mathcal{V}_a} \phi_x^N(x(\boldsymbol{\nu}_i) - x(\boldsymbol{\nu}_k))} \quad (2.19)$$

est un facteur de pondération compris entre 0 et 1. Il en découle une approximation de la dérivée de l'IM utilisée pendant la procédure d'optimisation. En posant $\mathbf{x}(\boldsymbol{\nu}_i) = (x^{(1)}(\boldsymbol{\nu}_i), x^{(2)}(\boldsymbol{\nu}_i))^T$ et en suivant la formule (2.9) pour la construction de l'IM, sa dérivée par rapport aux paramètres de la transformation $\boldsymbol{\mu}$ est donnée par :

$$\begin{aligned} \frac{d}{d\boldsymbol{\mu}} IM(\mathbf{x}) &= \frac{d}{d\boldsymbol{\mu}} H(x^{(2)}) - \frac{d}{d\boldsymbol{\mu}} H(\mathbf{x}) \\ &\approx \frac{1}{Card(\mathcal{V}_b)} \sum_{\boldsymbol{\nu}_j \in \mathcal{V}_b} \sum_{\boldsymbol{\nu}_i \in \mathcal{V}_a} (x^{(2)}(\boldsymbol{\nu}_j) - x^{(2)}(\boldsymbol{\nu}_i)) \\ &\quad \left[W_{x^{(2)}}(x^{(2)}(\boldsymbol{\nu}_j), x^{(2)}(\boldsymbol{\nu}_i)) \sigma_{x^{(2)}}^{-1} - W_{\mathbf{x}}(\mathbf{x}(\boldsymbol{\nu}_j), \mathbf{x}(\boldsymbol{\nu}_i)) \sigma_{x^{(2)}}^{-1} \right] \frac{d}{d\boldsymbol{\mu}} (x^{(2)}(\boldsymbol{\nu}_j) - x^{(2)}(\boldsymbol{\nu}_i)) \end{aligned} \quad (2.20)$$

avec

$$W_{x^{(2)}}(x^{(2)}(\boldsymbol{\nu}_j), x^{(2)}(\boldsymbol{\nu}_i)) = \frac{\phi_x^N(x^{(2)}(\boldsymbol{\nu}_j) - x^{(2)}(\boldsymbol{\nu}_i))}{\sum_{\boldsymbol{\nu}_k \in \mathcal{V}_a} \phi_x^N(x^{(2)}(\boldsymbol{\nu}_j) - x^{(2)}(\boldsymbol{\nu}_k))} \quad (2.21)$$

et

$$W_{\mathbf{x}}(\mathbf{x}(\boldsymbol{\nu}_j), \mathbf{x}(\boldsymbol{\nu}_i)) = \frac{\phi_{\mathbf{x}}^N(\mathbf{x}(\boldsymbol{\nu}_j) - \mathbf{x}(\boldsymbol{\nu}_i))}{\sum_{\boldsymbol{\nu}_k \in \mathcal{V}_a} \phi_{\mathbf{x}}^N(\mathbf{x}(\boldsymbol{\nu}_j) - \mathbf{x}(\boldsymbol{\nu}_k))} \quad (2.22)$$

Cette équation permet de conduire des recalages en utilisant un optimiseur de type descente de gradient. Dans l'article [Viola and Wells, 1997], le nombre de pixels des ensembles \mathcal{V}_a et \mathcal{V}_b donnés pour mener à bien la procédure est de cinquante. Cependant comme nous le verrons dans le chapitre 4, ce chiffre est sous-évalué. Et c'est un des problèmes de cette méthode : le coût calculatoire augmente de façon quadratique avec le nombre d'échantillons utilisés dans le calcul. Ainsi, en pratique, nous verrons qu'au delà de mille échantillons, un recalage d'images devient inenvisageable dans des temps de calculs raisonnables.

2.4.4 Estimateur à noyau B-Spline

Une autre méthode d'estimation très intéressante est à mettre à l'actif de [Thévenaz and Unser, 1997] (elle a été reprise notamment dans [Mattes et al., 2003]). Cette fois, le noyau utilisé est un noyau B-Spline, ce qui permet à la fois une implémentation simple et une unification conceptuelle et pratique entre la mesure de similarité, le modèle de transformation et l'interpolation.

Représentation des images et interpolation

Bien que nous abordions ici l'estimation des densités de probabilité, l'unification précédemment évoquée nous amène à préciser dès maintenant le modèle d'image utilisé dans [Mattes et al., 2003], puisque de ce modèle découle tout le reste de la démonstration. Nous donnerons les formules en 3 dimensions. En gardant les notations du chapitre 1, une image est décrite par un ensemble d'échantillons $x(\boldsymbol{\nu})$ définis sur une grille entière. Le calcul de $x(\boldsymbol{\nu})$ à des coordonnées non-entières requiert donc un schéma d'interpolation représentant l'image continue sous-jacente. Pour ce faire, une interpolation B-Spline est utilisée. En utilisant des coefficients que nous noterons c_i et qui sont calculés à partir des échantillons $x(\boldsymbol{\nu})$ de l'image en utilisant l'algorithme proposé par [Unser et al., 1991], les points de l'image se situant à des positions non-entières de la grille peuvent être interpolés par :

$$x_c(\boldsymbol{\nu}) = \sum_{i \in V} c_i \beta^{(3)}(x(\boldsymbol{\nu}) - x(\boldsymbol{\nu}_i)) \quad (2.23)$$

où $\beta^{(3)}$ est le noyau B-Spline de degré 3 présenté en annexe B. On dispose ainsi d'une représentation continue $x_c(\boldsymbol{\nu})$ de l'image. De la même manière, les auteurs présentent le gradient de l'image interpolée à n'importe quelle position :

$$\frac{\partial x_c(\boldsymbol{\nu})}{\partial \nu_x} = \sum_{i \in V} c_i \left(\frac{\partial \beta^{(3)}(x(\nu_x) - x(\nu_{x_i}))}{\partial \nu_x} \beta^{(3)}(x(\nu_y) - x(\nu_{y_i})) \beta^{(3)}(x(\nu_z) - x(\nu_{z_i})) \right) \quad (2.24)$$

De la même façon, on peut écrire les dérivées par rapport à ν_y et ν_z (ν_x, ν_y, ν_z étant les composantes spatiales 3D). Les dérivées explicites des B-Splines sont données en annexe B.

Information mutuelle

Une fois ce modèle d'image posé, nous pouvons nous intéresser à l'estimation de l'IM. A partir de la définition (2.11) l'opposé de l'IM en fonction du vecteur de paramètres $\boldsymbol{\mu}$ de la transformation considérée s'exprime par :

$$-IM(\mathbf{x}) = KL(p_{\mathbf{x}}(\mathbf{u}), p_{x^{(1)}}(u^{(1)})p_{x^{(2)}}(u^{(2)})) \quad (2.25)$$

avec

$$p_{\mathbf{x}}(\mathbf{u}) \approx \alpha \sum_{i \in V} \beta^{(0)}(u^{(1)} - x^{(1)}(\boldsymbol{\nu}_i)) \beta^{(3)}(u^{(2)} - x^{(2)}(\boldsymbol{\nu}_i)) \quad (2.26)$$

l'histogramme conjoint approximant la densité de probabilité conjointe. Dans cette expression, α est un facteur de normalisation qui assure que $\int p_{\mathbf{x}}(\mathbf{u}) d\mathbf{u} = 1$, $x^{(1)}(\boldsymbol{\nu}_i)$ et $x^{(2)}(\boldsymbol{\nu}_i)$ sont les valeurs normalisées d'intensité.

La densité marginale $p_{x^{(2)}}$ est estimée à partir de l'histogramme conjoint. Mais avant de présenter son expression, il faut savoir que le degré de la spline utilisé va déterminer la qualité de l'estimation [Thevenaz and Unser, 2000]. Le degré le plus bas possible $n = 0$ donne une estimation au plus proche voisin. Il en résulte des discontinuités dans l'histogramme estimé, et donc une optimisation plus délicate. Le degré suivant ($n = 1$) correspond à une interpolation linéaire. L'estimation est meilleure, et le calcul des coefficients B-Splines est trivial. Quand le degré des splines utilisées continue à augmenter, le crénelage de l'histogramme est bien réduit, mais au prix d'une mise en oeuvre plus coûteuse. Comme nous pouvons le voir dans l'expression de l'histogramme conjoint (2.26), un noyau spline de degré $n = 0$ (rectangle centré) est utilisé pour l'image de référence, alors que pour l'image flottante, un noyau de degré $n = 3$ (spline cubique) est utilisé. Cette différence vient du fait que l'image de référence ne contribue pas dans le calcul de la dérivée de la mesure de similarité puisqu'on dérive par rapport aux paramètres de la transformation et que l'image de référence n'est, par définition, par transformée. De ce fait, l'estimation de la densité de probabilité de l'image de référence n'a pas besoin d'être plus précise. A l'inverse, on aura besoin d'une densité de probabilité très lisse pour l'image flottante, d'où l'utilisation d'un noyau cubique. Ainsi, la densité marginale pour l'image flottante, calculée à partir de l'histogramme conjoint, est donnée par :

$$p_{x^{(2)}}(u^{(2)}) = \int_{\mathbb{R}} p_{\mathbf{x}}(u^{(1)}, u^{(2)}) du^{(1)} \quad (2.27)$$

L'histogramme marginal de l'image de référence peut être calculé indépendamment des paramètres de la transformation, en imposant à la fenêtre de Parzen B-Spline de satisfaire la partition de l'unité [Thevenaz and Unser, 2000]. Il s'écrit alors :

$$p_{x^{(1)}}(u^{(1)}) \approx \sum_{\boldsymbol{\nu}_i \in V^{(1)}} \beta^{(0)}(u^{(1)} - x^{(1)}(\boldsymbol{\nu}_i)) \quad (2.28)$$

Enfin, dans cet article [Thévenaz and Unser, 1997], le gradient de l'IM est également donné, ce qui permet en théorie une meilleure stratégie d'optimisation. En effet, comme nous l'avons vu au chapitre 1, la plupart des optimiseurs utilisés en recalage d'images se basent sur le sens de la dérivée, et utiliser une dérivée calculée de manière analytique plutôt qu'estimée empiriquement donne une solution plus précise. En dérivant la densité de probabilité conjointe (2.26) et en appliquant les règles de dérivation des fonctions composées, la i -ème dérivée partielle (c'est à dire la dérivée partielle par rapport au paramètre μ_i du vecteur $\boldsymbol{\mu}$ qui contient les différents paramètres de la transformation considérée) s'écrit :

$$\begin{aligned} \frac{\partial p_{\mathbf{x}}(\mathbf{u})}{\partial \mu_i} &\approx \frac{\alpha}{\Delta b_{FCard}(\mathcal{V})} \sum_{\boldsymbol{\nu}_i \in V} \beta^{(0)}(u^{(1)} - x^{(1)}(\boldsymbol{\nu}_i)) \\ &\times \frac{\partial \beta^{(3)}(u)}{\partial u} \Big|_{u=(u^{(2)}-x^{(2)}(\boldsymbol{\nu}_i))} \times \frac{-\partial x^{(2)}(\boldsymbol{\nu}_i)}{\partial \mu_i} \end{aligned} \quad (2.29)$$

Les formules pour les dérivées des splines sont données en annexe B. Le dernier terme de cette équation (2.29) correspond aux dérivées partielles de la transformation, et dépendra donc du type de transformation choisi. Dans ce modèle unifié où les déformations non-rigides sont également paramétrées par des B-Splines, une expression analytique de ce dernier terme peut être facilement dérivée.

L'avantage du noyau spline par rapport au noyau gaussien est le fait que le premier a un support fini, contrairement au second. Du point de vue du coût calculatoire, c'est intéressant puisque chaque échantillon (ou pixel) n'affecte qu'un faible nombre de classes. De plus, des méthodes récursives pour estimer les coefficients des B-Splines ont été proposées dans [Unser et al., 1991, Unser et al., 1993a, Unser et al., 1993b]. Elles permettent d'utiliser cet estimateur avec des coûts de calcul réduits afin d'envisager une utilisation du recalage d'images en routine clinique.

2.5 Limitations de l'IM et solutions existantes

Depuis 1995 et la parution de ces méthodes originelles d'utilisation de l'IM en recalage d'images, la popularité de cette mesure auprès de la communauté scientifique ne s'est jamais démentie. Toutefois, des critiques justifiées ont été formulées à son égard. Pour parer à ces limitations, de nombreux auteurs ont apporté des modifications ou des améliorations à l'IM. Savoir quelle méthode est à l'heure actuelle la plus efficace est une tâche impossible, puisque ces nouvelles méthodes sont souvent proposées pour une application spécifique et ne sont bien évidemment pas comparées à toutes les méthodes existantes. C'est pourquoi plutôt que d'essayer d'établir un classement des méthodes suivant leur efficacité, nous proposons de les classer suivant les grands axes de recherche couramment explorés pour contrer les faiblesses de l'IM. Nous ne pourrions pas être exhaustifs tant la littérature sur le sujet est abondante depuis les articles de [Collignon et al., 1995a, Viola, 1995]. De plus, la classification que nous proposons peut parfois paraître subjective, puisque certaines méthodes relèvent de plusieurs directions de recherche différentes. Dans ce cas, nous tenterons de classer les méthodes suivant les aspirations premières des auteurs. Ainsi, nous espérons que le lecteur trouvera dans ce chapitre les clés pour comprendre les récentes évolutions dans la mise en oeuvre de l'IM comme mesure de similarité en recalage d'images médicales. Nous donnerons enfin quelques pistes pour choisir telle ou telle métrique puisque quelques auteurs [Škerl et al., 2007, Cahill, 2010] ont réalisé des comparaisons entre les différentes implémentations actuelles de l'IM.

2.5.1 IM sensible au recouvrement partiel

L'IM est reconnue pour être sensible à la zone de recouvrement partiel des données. En effet, lorsque l'on recalc des images de modalités différentes, par exemple, il arrive qu'une structure, ou plus généralement qu'une partie des données présente dans une des images ne le soit pas dans l'autre. C'est la notion de recouvrement partiel. De plus, au cours du recalage, la transformation géométrique appliquée à l'image flottante peut modifier le volume de recouvrement des deux images.

Plusieurs critères ont été développés pour répondre à cette problématique. Le plus connu est sans doute l'**Information Mutuelle Normalisée** (IMN) [Studholme et al., 1997], apparue peu après les premières méthodes de recalage utilisant l'IM comme mesure de similarité. Celle-ci consiste simplement en un réordonnancement des différentes entropies composant l'IM :

$$IMN(\mathbf{x}) = \frac{H(x^{(1)}) + H(x^{(2)})}{H(\mathbf{x})} \quad (2.30)$$

Cette mesure est souvent préférée à l'IM en recalage d'images, car sa robustesse au recouvrement partiel est effectivement supérieure à celle de l'IM. Elle a d'ailleurs été largement utilisée en recalage d'images médicales, et le lecteur pourra trouver de nombreuses références de son utilisation dans [Pluim et al., 2003]. Quel que soit l'estimateur de l'IM utilisé, il est possible de se ramener à sa version normalisée. Cela peut poser quelques problèmes quand on étudie la littérature. En effet, les auteurs parlent par exemple souvent d'IMN sans préciser quel est l'estimateur sous-jacent. Implicitement, on pourrait considérer qu'il s'agit de l'estimateur utilisé par [Studholme et al., 1997], c'est à dire basé sur l'estimateur à histogramme des densités de probabilité, puisque c'est lui qui a proposé l'IMN. Cependant, l'IMN est parfois calculée à partir de l'estimateur à noyau B-Spline de [Thévenaz and Unser, 1997], puisque cette méthode d'estimation est plus efficace comme nous le verrons par la suite. Aussi, quand nous parlerons d'IMN, nous préciserons quand c'est nécessaire l'estimateur utilisé.

Deux autres mesures de similarité ont été proposées par [Maes et al., 1997] pour répondre à ce problème de recouvrement partiel. Il s'agit du **Rapport de Corrélation Entropique** (RCE ou Entropy Correlation Coefficient), défini par :

$$\begin{aligned} RCE(\mathbf{x}) &= \sqrt{2 - \frac{2H(\mathbf{x})}{H(x^{(1)}) + H(x^{(2)})}} \\ &= \frac{2IM(\mathbf{x})}{H(x^{(1)}) + H(x^{(2)})} \end{aligned} \quad (2.31)$$

et de l'**information exclusive** qui s'exprime par :

$$IE(\mathbf{x}) = H(\mathbf{x}) - IM(\mathbf{x}) = 2H(\mathbf{x}) - H(x^{(1)}) - H(x^{(2)}) \quad (2.32)$$

Plus récemment [Cahill et al., 2008], les mesures précédemment évoquées ont été mises en défaut sur un problème de recouvrement partiel simple. Les auteurs ont ainsi montré que même l'IMN et le RCE étaient sensibles au recouvrement partiel, et ils ont défini de nouvelles mesures, nommées **Information Mutuelle Modifiée** (IMM ou MMI pour Modified Mutual Information), **Information Mutuelle Normalisée Modifiée** (IMNM ou MNMI pour Modified Normalized Mutual Information) et **Rapport de Corrélation Entropique Modifié** (RCEM ou MECC pour Modified Entropy Correlation Coefficient), qui se prévalent à leur sens de ce problème, du moins sur les exemples présentés. Pour ce faire, les statistiques (moyennes et écart-type) des images de référence et flottantes doivent être calculées avant le recalage et sont incorporées dans les mesures originelles. Nous ne rentrerons pas dans les détails de cette étude car bien qu'intéressante, elle ne remet sans doute pas en cause l'efficacité avérée de l'IMN classique dans de nombreux problèmes de recalage. En effet, même plus de 10 ans après sa publication, elle reste toujours une mesure de référence en recalage d'images médicales multimodales. Dans la majorité des améliorations apportées à l'IM que nous présentons dans ce qui suit, elle reste toujours la mesure de comparaison qui est donnée pour évaluer les nouvelles méthodes.

2.5.2 IM incorporant une information spatiale

Un désagrément reconnu de l'IM est qu'elle ne fait aucune supposition sur l'information spatiale contenue dans les images à recaler [Rueckert et al., 2000]. Traitant les pixels/voxels des images comme des réalisations d'une VA, la localisation spatiale de ces pixels n'entre pas du tout en compte dans l'estimation de l'IM. Dans le cadre du recalage d'images médicales, il peut sembler intéressant, par exemple, de se servir des informations communes telles que les structures anatomiques présentes dans les deux images pour améliorer la robustesse de l'estimation. Cette information supplémentaire ferait se rapprocher l'IM des techniques de recalage géométrique présentées succinctement au chapitre 1. Un exemple très simple tiré de [Rueckert et al., 2000] permet d'illustrer ce problème. Les 2 images données sur la figure 2.4 ont la même entropie marginale alors qu'elles sont clairement différentes et que pour n'importe quel observateur humain, seule l'image de gauche apporte une information. En fait, l'image de droite a été obtenue en déplaçant aléatoirement les pixels de l'image de gauche, et le calcul de l'entropie marginale donne le même résultat dans les deux cas.

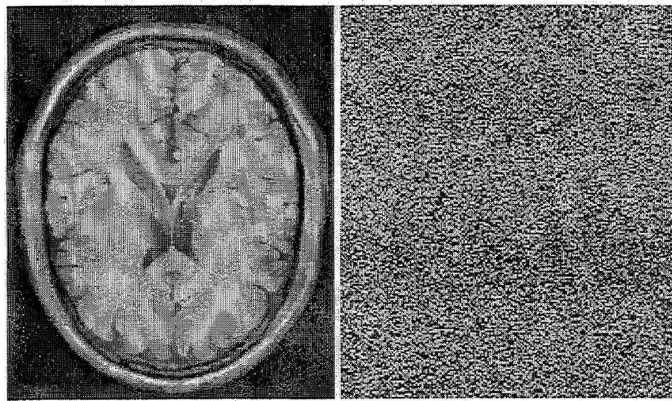


FIGURE 2.4 – L'IM classique ignore l'information spatiale. Les 2 images ont la même entropie marginale, l'image de droite étant créée en déplaçant aléatoirement les pixels de l'image de gauche. D'après [Rueckert et al., 2000].

Une autre raison poussant à intégrer de l'information spatiale dans l'IM est son utilisation pour le recalage non-rigide. En effet comme nous l'avons vu précédemment, le calcul de l'IM est en général basé sur un histogramme conjoint global. Pendant l'optimisation, l'algorithme peut avoir tendance à minimiser des minima locaux plutôt que le minimum global. Avec un modèle de transformation affine, ce problème ne se pose généralement pas trop, du fait du faible nombre de degrés de liberté (DDL). Mais lorsque l'on passe à un recalage non-rigide, le nombre important de paramètres à optimiser peut entraîner une divergence de l'algorithme.

Pour remédier à ces difficultés, plusieurs auteurs ont tenté de prendre en compte cette information issue des images, en utilisant l'emplacement des pixels ou les relations pouvant exister entre pixels voisins. Nous allons maintenant détailler quelques unes de ces approches, les plus intéressantes aux yeux de la communauté du recalage iconique.

IM incorporant une information de gradient

Dans [Pluim et al., 2000], les auteurs proposent d'utiliser non seulement les valeurs des niveaux de gris dans l'image, mais également les gradients spatiaux de ces images dans le calcul de l'IM. En effet, de forts gradients indiquent en général une transition entre des tissus de nature

différente dans les images médicales, et sont donc porteurs d'une information importante. Pour ce faire, un vecteur de gradients est construit pour chaque point $\boldsymbol{\nu}^{(1)}$ de l'image de référence et son point correspondant $\boldsymbol{\nu}^{(2)}$ dans l'image flottante, qui est trouvé après transformation géométrique de l'image de référence. Les trois (pour des images tri-dimensionnelles) dérivées directionnelles sont calculées en utilisant les dérivées premières d'un noyau gaussien d'écart type σ . L'angle entre les vecteurs de gradients est ensuite défini par :

$$\alpha_{\boldsymbol{\nu}^{(1)}, \boldsymbol{\nu}^{(2)}}(\sigma) = \arccos \left(\frac{\nabla_{\boldsymbol{\nu}^{(1)}}(\sigma)^T \nabla_{\boldsymbol{\nu}^{(2)}}(\sigma)}{\|\nabla_{\boldsymbol{\nu}^{(1)}}(\sigma)\| \|\nabla_{\boldsymbol{\nu}^{(2)}}(\sigma)\|} \right) \quad (2.33)$$

avec $\nabla_{\boldsymbol{\nu}}(\sigma)$ le vecteur gradient colonne d'écart type σ . Quelques manipulations mathématiques sont ensuite nécessaires. Tout d'abord, il faut favoriser les angles très petits ou proches de π , pour contrer le fait que les structures anatomiques dans des images multimodales peuvent avoir des gradients de même orientation mais de directions opposées. C'est le rôle de la fonction de pondération $w(\alpha)$ de l'équation (2.36). Ensuite, seuls les gradients forts apparaissant dans les deux images sont pris en compte, puisqu'ils traduisent des structures communes. Finalement, le terme (2.33), modifié pour prendre en compte les caractéristiques précédemment évoquées, est construite pour tous les échantillons des deux images, donnant naissance à un terme de gradient $G(\mathbf{x})$ qui est incorporé dans la mesure d'IM :

$$IM_{new}(\mathbf{x}) = G(\mathbf{x})IM(\mathbf{x}) \quad (2.34)$$

avec :

$$G(\mathbf{x}) = \sum_{(\boldsymbol{\nu}^{(1)}, \boldsymbol{\nu}^{(2)}) \in D^{(1)} \times D^{(2)}} w(\alpha_{\boldsymbol{\nu}^{(1)}, \boldsymbol{\nu}^{(2)}}(\sigma)) \min(\|\nabla_{\boldsymbol{\nu}^{(1)}}(\sigma)\|, \|\nabla_{\boldsymbol{\nu}^{(2)}}(\sigma)\|) \quad (2.35)$$

et

$$w(\alpha) = \frac{\cos(2\alpha) + 1}{2} \quad (2.36)$$

Les expérimentations menées montrent une amélioration relative des résultats en recalage en utilisant cette nouvelle mesure plutôt que l'IM classique. Le gain apporté par cette mesure est une meilleure robustesse. L'ajout d'information spatiale tend en effet à réduire le nombre de minima locaux de la fonction de coût. Cette nouvelle mesure ne peut-être vue comme une extension de l'IM, puisqu'elle consiste simplement à multiplier la mesure par un terme prenant en compte l'information de gradient. Elle fait presque figure d'exception dans les méthodes proposant l'ajout d'information spatiale, car elle cherche à garder la définition initiale de l'IM. La majorité des auteurs ont fait le choix au contraire de modifier la définition en ajoutant un terme d'attache à l'information spatiale. Ce sont ces méthodes que nous présentons maintenant.

IM de dimension supérieure

Incorporer une information supplémentaire directement dans l'IM n'est pas simple, car dans la plupart des cas, cela se traduit par la définition d'une IM d'un couple de deux vecteurs aléatoires de dimension 2. Nous entendons par là que plutôt de se contenter d'estimer l'IM entre une image flottante et une image de référence, l'estimation se fait entre ces deux images et entre des images ou vecteurs de caractéristiques supplémentaires. Nous présentons quelques unes de ces méthodes, et leurs limitations essentiellement dues à la construction d'un histogramme conjoint de dimension supérieure à deux.

Dans [Rueckert et al., 2000], qui est sans doute le premier article proposant d'incorporer une information spatiale à l'IM, et qui arrive quelques années seulement après les articles fondateurs, il est suggéré d'utiliser à la fois l'intensité, mais aussi l'intensité du voisinage pour

construire l'IM. Pour cela, les auteurs définissent des **mesures d'information du second ordre**, à savoir les **entropies marginales et conjointe du second ordre** définies respectivement par :

$$H_2(x^{(i)}) = - \oint_{\mathbb{R}^2} p_{x^{(i)}}(u_1^{(i)}, u_2^{(i)}) \log(p_{x^{(i)}}(u_1^{(i)}, u_2^{(i)})) du_1^{(i)} du_2^{(i)} \quad (2.37)$$

et

$$H_2(\mathbf{x}) = - \oint_{\mathbb{R}^4} p_{\mathbf{x}}(\mathbf{u}_1, \mathbf{u}_2) \log(p_{\mathbf{x}}(\mathbf{u}_1, \mathbf{u}_2)) d\mathbf{u}_1 d\mathbf{u}_2 \quad (2.38)$$

Dans l'équation (2.37), $p_{x^{(i)}}$ correspond à la probabilité qu'un voxel ν_1 de l'image $x^{(i)}(\nu)$ ait une intensité $u_1^{(i)}$ et son voxel voisin ν_2 une intensité $u_2^{(i)}$. Cette probabilité jointe peut être estimée à partir d'un histogramme conjoint qui ne mesure plus cette fois-ci les co-occurrences des intensités entre paires de voxels dans des images différentes, mais entre des voxels voisins dans la même image. De la même manière, l'entropie conjointe de second ordre (2.38) est calculée entre non seulement les paires d'intensités provenant des deux images à recaler, mais également de leur voisin, ce qui nécessite la construction d'un histogramme 4D. L'interprétation qui peut être faite de ces nouveaux histogrammes conjoints est relativement simple : les voxels situés dans une région homogène contribueront à la diagonale de l'histogramme conjoint, tandis que les voxels aux frontières entre deux zones d'intensités différentes (typiquement entre des structures anatomiques différentes) feront s'étaler l'histogramme conjoint. Ensuite, les mesures classiques que sont la MI et la NMI sont réécrites avec ces nouvelles définitions de l'entropie.

Deux autres approches intéressantes et comparables sont à mettre à l'actif de [Studholme et al., 2006] et [Loeckx et al., 2009]. Dans le premier article, l'IM intègre une information locale des images à recaler. La mesure IMR (pour **IM Régionale**) permet ainsi de prendre en compte les variations locales d'intensités. Pour ce faire, l'IM d'un vecteur aléatoire de dimension 3 est estimée. Notre vecteur $\mathbf{x} = (x^{(1)}, x^{(2)})^\top$ est étendu avec une troisième composante représentant la position spatiale. Nous noterons $x^{(3)}$ la VA associée à cette information spatiale. Dans cette formulation, nous notons donc $\mathbf{x} = (x^{(1)}, x^{(2)}, x^{(3)})^\top$ ce vecteur aléatoire étendu. L'expression de l'IMR est alors :

$$IMR(\mathbf{x}) = H(x^{(1)}) + H(x^{(2)}) + H(x^{(3)}) - H(\mathbf{x}) \quad (2.39)$$

De manière similaire, le second article [Loeckx et al., 2009] proposent l'IM Conditionnelle (IMC) qui utilise également une troisième VA représentant la distribution spatiale des intensités conjointes. Leur formulation peut s'écrire :

$$IMC(\mathbf{x}) = H(x^{(1)}|x^{(3)}) + H(x^{(2)}|x^{(3)}) - H(x^{(1)}, x^{(2)}|x^{(3)})$$

On voit bien à travers ces exemples la volonté d'incorporer davantage d'information spatiale dans l'IM. Ils sont en outre représentatifs des avancées récentes : construire un vecteur de caractéristiques de dimension supérieure et estimer les entropies marginales et conjointes associées. Malheureusement se pose alors la question de la construction d'histogrammes conjoints de dimension supérieure à 2. C'est une faiblesse de ces méthodes, puisque la construction de ces histogrammes multidimensionnels s'avère délicate, comme nous l'expliquerons dans la section 2.5.3. Si les articles précédents ont le mérite de montrer le problème lié au manque d'information spatiale apportée par l'IM, la réponse apportée, néanmoins, est en pratique très compliquée à mettre en oeuvre. D'autres articles, tels que [Holden et al., 2004] se sont également heurtés à ce problème de construction d'un histogramme conjoint de dimension supérieure.

Récemment, des approximations mathématiques permettant de définir une IM multidimensionnelle ont été proposées, et nous les approfondirons dans la section suivante 2.5.3. A partir de ces nouvelles méthodes et notamment des graphes d'entropie, le calcul d'entropies multidimensionnelles s'est vu simplifié, et certains auteurs en ont tiré partie. Ainsi, dans [Sabuncu and Ramadge, 2003], l'utilisation de l'information spatiale contenue dans l'image est proposée. Au lieu de calculer l'IM simplement entre les niveaux de gris, des vecteurs de caractéristiques spatiales sont utilisés. Dans ce même contexte des graphes d'entropie, l'article de [Staring et al., 2009] a récemment défini l'IM multi-caractéristiques (MultiFeatures Mutual Information). Dans cet article, encore une fois, les caractéristiques de l'image prises en compte doivent apporter une connaissance supplémentaire sur la structure locale ou spatiale des images. Les auteurs ont fait le choix des caractéristiques cartésiennes invariantes de l'image jusqu'à l'ordre 2. Derrière cette dénomination se cachent la luminance L , les gradients spatiaux $\mathbf{g} = \nabla_{\mathbf{v}} L$ et le hessien \mathbf{H} de L . La liste de ces caractéristiques est donnée dans la table 2.1, où $tr(\cdot)$ représente la trace de la matrice. Ces caractéristiques sont invariantes par rotation et translation et irréductibles. Elles sont placées dans un vecteur de caractéristique et l'IM est calculée grâce au graphe des k plus proches voisins (kNN graph). Cette méthode montre des résultats intéressants, même si le coût calculatoire reste important.

Notation d'Einstein	Notation matricielle
L	L
$L_i L_i$	$\mathbf{g}^T \mathbf{g}$
$L_i L_{ij} L_j$	$\mathbf{g}^T \mathbf{H} \mathbf{g}$
$L_i L_{ij} L_{jk} L_k$	$\mathbf{g}^T \mathbf{H} \mathbf{H} \mathbf{g}$
L_{ii}	$tr(\mathbf{H})$
$L_{ij} L_{ji}$	$tr(\mathbf{H} \mathbf{H})$
$L_{ij} L_{jk} L_{ki}$	$tr(\mathbf{H} \mathbf{H} \mathbf{H})$

TABLE 2.1 – Caractéristiques prises en compte pour le calcul de l'IM multi-caractéristiques, d'après [Staring et al., 2009].

2.5.3 IM se prémunissant du calcul de l'histogramme conjoint

Quelque soit la méthode d'estimation de la densité de probabilité choisie, les paramètres de lissage, tout comme le nombre de classes de l'histogramme, ou la largeur des noyaux pour les estimateurs de type Parzen, doivent être choisis par l'utilisateur. Les performances des méthodes d'estimation dépendent grandement de ces paramètres, et ce faisant un mauvais choix initial peut aboutir à des résultats désastreux. Pour s'en convaincre, il suffit de regarder l'exemple synthétique de la figure 2.5 qui donne 3 histogrammes monodimensionnels traduisant le même jeu de données (une loi normale centrée-réduite). On voit bien que les paramètres choisis influent énormément sur la forme finale de l'histogramme et donc sur l'estimation de la densité de probabilité sous-jacente. Même quand un estimateur à noyau est utilisé, des phénomènes de sur-lissage ou de sous-lissage peuvent apparaître en fonction du choix du noyau, comme le montre la figure 2.6.

De plus, l'estimation de l'histogramme conjoint entraîne une charge calculatoire très importante : 90% du temps, pendant un processus de recalage, est ainsi consacré à l'estimation de cet histogramme [Sarrut, 2000]. Pour réduire cette charge, il est tentant de réduire le nombre d'échantillons (de pixels de l'image) utilisés pour le calcul de cet histogramme. Cependant, la fiabilité de l'estimation par l'histogramme conjoint repose entièrement sur ce nombre d'échan-

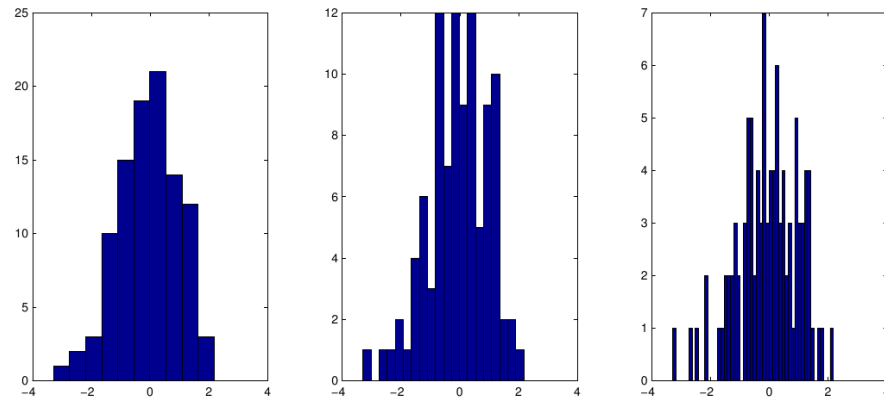


FIGURE 2.5 – Ces trois histogrammes ont été générés avec les mêmes données mais avec un nombre de classes, et une largeur de ces classes différente.

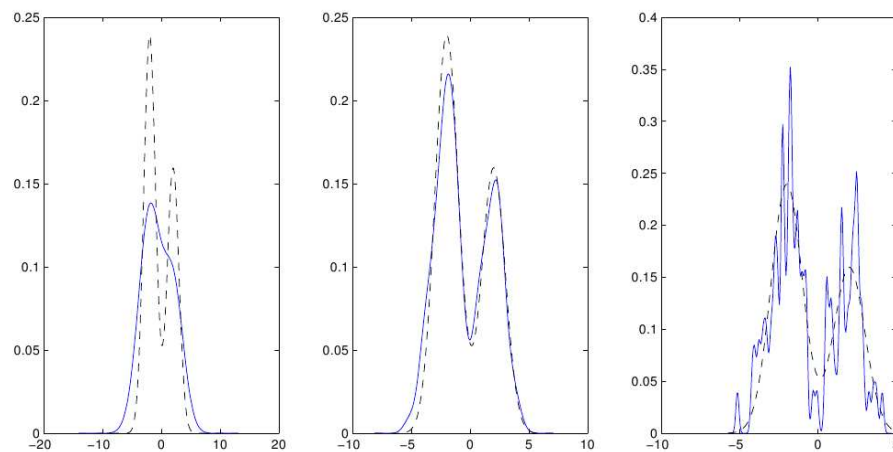


FIGURE 2.6 – Influence du paramètre de lissage dans les estimateurs à noyau. A gauche, l'histogramme est sur-lissé, tandis qu'à droite, il est sous lissé (la vraie densité est donnée en pointillés).

tillons utilisés pour son estimation, et un nombre trop faible entraîne un biais d'estimation.

Enfin, outre le choix des paramètres de cet histogramme, l'extension de l'histogramme conjoint aux dimensions supérieures à 2, qui est à l'heure actuelle une piste importante pour apporter des informations supplémentaires à l'IM comme discuté dans la section 2.5.2, n'est pas du tout trivial. En effet, cela entraîne :

- un problème de paramétrisation (choix du nombre de classes, du facteur de lissage, etc.) : ce problème, déjà délicat pour des histogrammes bidimensionnels, le devient encore plus pour des dimensions supérieures.
- une charge calculatoire très importante. Si l'on ne prend ne serait-ce que 64 classes par dimension, un histogramme à 4 dimensions comme utilisé dans [Rueckert et al., 2000] a déjà plus de 16 millions de classes. De plus, à nombre de classes constant dans l'histogramme, le nombre moyen d'échantillons par classe décroît exponentiellement en fonction de la dimension de l'histogramme conjoint. Ainsi, si l'on veut garder constant le nombre moyen d'échantillons par classe donnée, le nombre de classes doit décroître exponentiellement en fonction de la dimension de l'histogramme conjoint. Pour s'en convaincre, nous pouvons nous intéresser à l'exemple donné par [Russakoff et al., 2004], repris figure 2.7 et qui illustre le problème lié aux dimensions supérieures dans la construction de l'histogramme conjoint. Ce graphe donne, sur un exemple synthétique, la manière dont l'entropie estimée va s'approcher de l'entropie réelle, pour des histogrammes 4-D, 5-D et 6-D, en fonction du nombre d'échantillons utilisés. Nous pouvons remarquer que même en utilisant 2 millions d'échantillons (ce qui dans le cadre du recalage implique des temps de calcul énormes), l'estimation en 6-D reste approximative (moins de 80% de la vraie valeur).

Face à tous ces problèmes, l'idée de se prémunir de la construction de l'histogramme conjoint pour calculer l'IM a rencontré un succès important ces dernières années. Nous ne rentrerons pas dans le détail de toutes les méthodes qui sont actuellement développées, car les approches sont variées et ne rencontrent pas toutes le même succès. Mais il nous semble important de donner un panorama des recherches actuelles sur le sujet.

Les travaux [Cahill et al., 2006] proposent une méthode permettant d'incorporer de l'information spatiale dans l'IM sans avoir recours aux histogrammes N -dimensionnels. Pour cela, ils partitionnent l'espace N -dimensionnel en un ensemble de régions et approximent l'entropie en formant un histogramme mono-dimensionnel avec cet ensemble de régions. Les résultats présentés ne permettent cependant pas de se faire une réelle idée de l'efficacité de cette méthode en recalage d'images médicales.

Dans [Russakoff et al., 2004], le travail de [Rueckert et al., 2000] est étendu à des voisinages plus importants. Pour se prémunir de la construction d'histogrammes conjoints de trop grandes dimensions, les auteurs parviennent à diviser le calcul de l'entropie d'une distribution de probabilité N -dimensionnelle en la somme de N distributions mono-dimensionnelles. Pour cela, ils utilisent les propriétés mathématiques de l'entropie d'une distribution discrète (invariance en rotation et translation) et se servent de l'Analyse en Composantes Principales (ACP) pour décorréler les VAs mises en jeu. Ce travail est intéressant puisque leur nouvelle mesure, l'**IM Régionale** (IMR), qui est seulement une homonyme de l'IMR de [Studholme et al., 2006], est donnée comme plus robuste que l'IM traditionnelle. Cependant, le calcul de cette variante de l'IM nécessite d'importantes hypothèses comme la décorrélation des VAs mises en jeu et leur gaussianité.

Le travail [De Nigris et al., 2010] qui s'appuie sur les articles de [Karaçali, 2007], [Biesdorf et al., 2009] est très intéressant, puisque une fois encore, il se propose de se prémunir de l'estimation des densités de probabilité par histogramme conjoint. Pour ce faire, [Karaçali, 2007] décrit une mesure de l'IM calculable sur un voisinage très restreint. L'idée

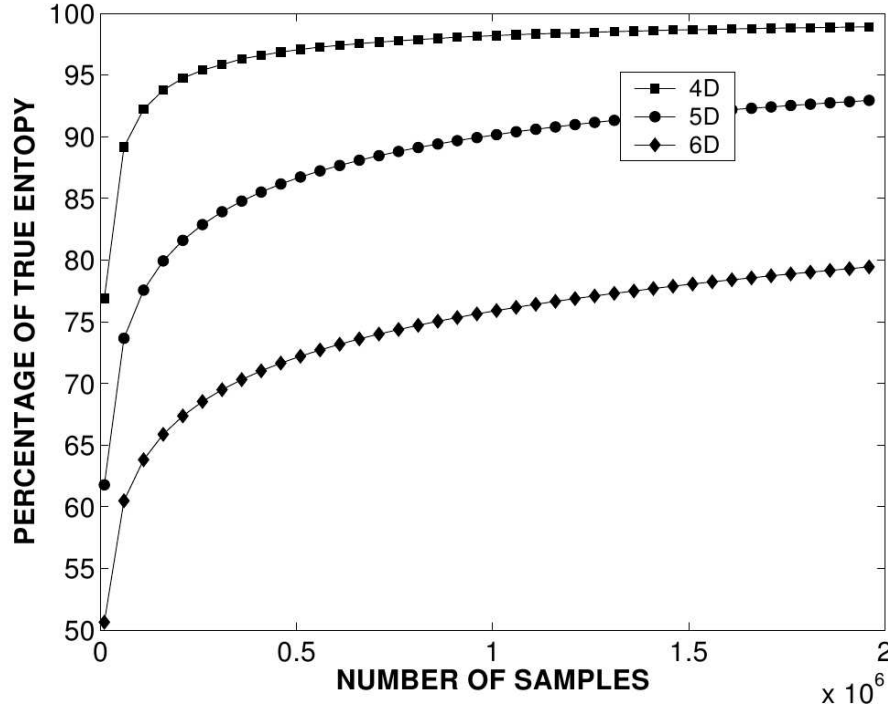


FIGURE 2.7 – Illustration du problème d'estimation par histogramme conjoint pour les dimensions supérieures extrait de [Russakoff et al., 2004]. Graphique donnant le pourcentage par rapport à l'entropie vraie en fonction du nombre d'échantillons utilisés pour des histogrammes 4-D, 5-D et 6-D. Quand la dimension de l'histogramme croît linéairement, il faut un nombre exponentiellement croissant d'échantillons pour estimer correctement l'entropie.

pour calculer cette **IM Locale** (IML) est de modéliser les images par un développement de Taylor du premier ordre qui suit une décroissance monotone en fonction de l'angle θ entre les orientations des gradients.

L'alternative qui rencontre le plus de succès à l'heure actuelle est sans doute l'estimation par graphes d'entropie développée dans [Ma et al., 2000, Hero et al., 2002]. De nombreux auteurs utilisent cet estimateur dans le cadre du recalage multimodal d'images médicales [Ma et al., 2000, Sabuncu and Ramadge, 2003, Neemuchwala and Hero, 2005, Kybic, 2007], [Sabuncu and Ramadge, 2008, Staring et al., 2009]. La formulation de cette méthode d'estimation ne se base plus sur l'entropie de Shannon (2.2) mais sur l'**entropie de Renyi** (2.43) qui sera développée dans la section suivante 2.5.4, équation (2.43). Nous faisons donc ici une petite digression dans la progression du chapitre. Ceci vient du fait que les graphes d'entropies répondent à 2 problématiques à la fois : se prémunir du calcul de l'histogramme conjoint, tout en utilisant une autre mesure de l'entropie, et nous pensons que cette méthode est plus importante du point de vue de la première problématique que de la seconde, qui est juste une conséquence de la première.

Plusieurs familles de graphes d'entropie se distinguent dans la manière de construire le graphe. Nous présenterons ici la méthodologie générale pour contruire un graphe d'entropie, mais le lecteur trouvera par exemple dans [Sabuncu, 2006] les différents types de graphes utilisés en recalage. De manière générale, un graphe $G(A, S)$ est composé d'un nombre fini de sommets $S \subset \mathbb{R}^2$ et d'arêtes A (cf. figure 2.8 pour un exemple synthétique). Chaque arête $\mathbf{a} = (\mathbf{s}_1, \mathbf{s}_2) \in A$ possède une longueur Euclidienne $\|\mathbf{a}\| = \|\mathbf{s}_1 - \mathbf{s}_2\|$. Pour $\gamma \in \mathbb{R}$, le γ -poids

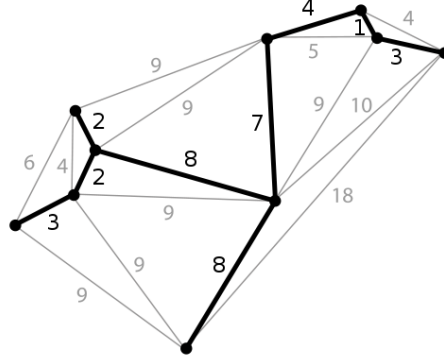


FIGURE 2.8 – Exemple synthétique de graphe à chemin minimal. Un poids est attribué à chaque arête reliant 2 sommets. Ici, les poids sont grossièrement proportionnels à la longueur des arêtes.

de G est donné par :

$$W_\gamma(G) = \sum_{\mathbf{a} \in \mathcal{A}} \|\mathbf{a}\|^\gamma \quad (2.40)$$

Dans la suite, nous ne considérons qu'une certaine classe de graphes $\mathcal{G}_C(\mathcal{S})$ se conformant à une contrainte topologique C et ayant le même jeu de sommets \mathcal{S} . $\mathcal{G}_C(\mathcal{S})$ inclut notamment les arbres couvrant minimaux (Minimum Spanning Trees ou MST), et les graphes des k voisins (k -Neighbor graphs ou k -N graphs) qui sont les plus utilisés dans le contexte du recalage. Dans la suite, nous faisons l'hypothèse que C est fixé.

Pour une famille de graphes $\mathcal{G}(\mathcal{S})$, nous définissons le γ -poids minimum de $\mathcal{G}(\mathcal{S})$ comme :

$$W_\gamma^*(\mathcal{S}) = \min_{G \in \mathcal{G}(\mathcal{S})} W_\gamma(G) \quad (2.41)$$

et le graphe de poids minimal $G^*(\mathcal{S}) = \arg \min_{G \in \mathcal{G}(\mathcal{S})} W_\gamma(G)$, qui n'est pas forcément unique.

Ces graphes sont intéressants car ils permettent de construire une estimée de l' α -entropie, ou entropie de Renyi. Considérons le vecteur aléatoire \mathbf{s} prenant ses valeurs dans $[0, 1]^2$, de densité de probabilité de Lebesgue $p_{\mathbf{s}}$ et $\mathcal{S} = \{\mathbf{s}_1, \dots, \mathbf{s}_N\}$ est un ensemble d'échantillons indépendants de \mathbf{s} . Posons $\gamma = 2(1 - \alpha)$ et :

$$\tilde{H}_\alpha(\mathbf{s}) = \frac{1}{1 - \alpha} \log \left(\frac{W_{2(1-\alpha)}^*(\mathcal{S})}{N^\alpha} \right) \quad (2.42)$$

Il est montré dans [Hero and Michel, 1999] que, $\forall \alpha \in (0, 1)$, $\tilde{H}_\alpha(\mathcal{S}) + \log \beta / (1 - \alpha)$ est un estimateur asymptotiquement non-biaisé et consistant de l'entropie de Renyi (2.43), où β est une constante dépendant de la contrainte topologique C et du paramètre α , mais pas de $p_{\mathbf{s}}$. Ainsi, à partir de (2.42), il est possible de construire une approximation de l' α -IM, l'IM basée sur l'**entropie de Renyi**, comme l'ont fait par exemple [Neemuchwala and Hero, 2005, Staring et al., 2009].

Cette méthode semble prometteuse dans la mesure où elle permet d'approximer l'entropie de Renyi sans fixer de paramètres, et dans des espaces multidimensionnels, la complexité en terme de temps de calcul croissant linéairement en fonction du nombre N de dimensions.

Cependant, des efforts doivent être consentis pour augmenter la rapidité de recherche du chemin minimal dans ces graphes.

2.5.4 IM basées sur une autre mesure de l'entropie

Depuis l'apparition de l'entropie de Shannon [Shannon, 1948], de nombreuses mesures d'entropie ont émergé dans différents contextes applicatifs. La définition originelle de l'IM en recalage se base sur l'entropie de Shannon, mais il est possible de construire de nouvelles définitions basées sur des entropies différentes. Ici, les auteurs ne cherchent pas forcément à corriger les défauts inhérents à l'IM mais plutôt à définir des mesures d'IM plus générales. En effet, nombre de ces mesures d'entropie sont en fait des généralisations de l'entropie de Shannon ou entropie différentielle (2.2). Parmi celles-ci, on compte notamment l'entropie de Renyi [Renyi, 1960]

Entropie de Renyi

Elle se définit de la façon suivante :

$$H_\alpha(x) = \frac{1}{1-\alpha} \log \left(\int_{\mathbb{R}} p_x^\alpha(u) du, \alpha > 0, \alpha \neq 1 \right) \quad (2.43)$$

qui est l'une des plus usitées et qui converge vers l'entropie de Shannon lorsque $\alpha \rightarrow 1$. En recalage d'images, plusieurs auteurs dont [Pluim et al., 2001, Wachowiak et al., 2003] se sont intéressés à cette entropie pour calculer une mesure d'IM notée IM_α . C'est une mesure au sens où elle permet de mesurer l'écart entre la densité de probabilité conjointe p_x et le produit des deux densités marginales $p_{x(1)}p_{x(2)}$.

$$IM_\alpha(x) = \frac{1}{\alpha(\alpha-1)} \left(\oint_{\mathbb{R}^2} \frac{p_x(u)^\alpha}{(p_{x(1)}(u^{(1)})p_{x(2)}(u^{(2)}))^{\alpha-1}} du - 1 \right), \alpha \neq 0, \alpha \neq 1 \quad (2.44)$$

Quand $\alpha \rightarrow 1$, IM_α est équivalent à la définition de l'IM donnée par (2.11). Cette mesure est intéressante en elle même, mais c'est surtout l'entropie de Renyi qui a intéressé les chercheurs puisqu'elle permet, outre ses qualités intrinsèques, de calculer la densité de probabilité conjointe sans passer par l'histogramme conjoint, en utilisant les graphes d'entropie que nous avons présentés dans le paragraphe 2.5.3. Dans ce cadre, il a été montré [Ma et al., 2000] que l'entropie de Renyi quadratique ((2.43) avec $\alpha = 2$) pouvait être utilisée en recalage d'images.

Il faut noter que l'IM basée sur l'entropie de Rényi IM_α n'est qu'un cas particulier d'une classe plus générale de mesures appelées f -mesures [Pluim et al., 2001], qui font partie des f -divergences et sont données par :

$$f(x) = \oint_{\mathbb{R}^2} (p_{x(1)}(u^{(1)})p_{x(2)}(u^{(2)})) f \left(\frac{p_x(u)^\alpha}{p_{x(1)}(u^{(1)})p_{x(2)}(u^{(2)})} \right) du \quad (2.45)$$

parmi lesquelles on compte notamment la V -information ou la M_α -information [Vajda, 1989], mais qui n'entrent pas dans la problématique de ce chapitre, puisqu'elles ne permettent pas de définir l'IM.

Nous citons brièvement ici une autre mesure d'entropie, l'entropie de Havrda-Charvat de type q , donnée par :

$$H_q(x) = \frac{1}{1-q} \int_{\mathbb{R}} p_x^q(u)(1-p_x^{1-q}(u))du, q \in \mathbb{R} \setminus \{0, 1\} \quad (2.46)$$

qui a également été utilisée pour construire une IM en recalage d'images médicales [Wachowiak et al., 2003], mais qui n'a pas la même reconnaissance que l' IM_α dans la littérature.

Entropie généralisée

L'estimation des densités de probabilité n'est pas un préalable indispensable à l'estimation de l'entropie. Pour preuve, [Liao and Chung, 2006] propose l'utilisation de la fonction de répartition comme alternative crédible dans le cadre du recalage multimodal d'images médicales. Cette approche, proposée initialement par [Wang et al., 2003, Rao et al., 2004], trouve son crédit mathématiquement : la première justification à cette approche est que la fonction de répartition existe même dans les cas où les densités de probabilité ne sont pas définies.

De plus, lorsque la densité de probabilité est définie, la fonction de répartition est plus régulière que la densité de probabilité puisque cette dernière est sa dérivée. Enfin, la définition originelle de l'entropie de Shannon était donnée pour des VAs discrètes, mais sa définition continue (2.2) n'est pas une conséquence directe du cas discret. Pour des distributions continues, l'entropie de Shannon peut ainsi être négative ou infinie.

Pour décrire cette méthode d'estimation de l'entropie, nous nous appuyons sur l'article [Zografos and Nadarajah, 2005] où sont présentées l'**Entropie Survivante Exponentielle** (ESE) et l'**Entropie Survivante Exponentielle Généralisée** (ESEG) qui généralisent l'entropie cumulative résiduelle (ECR) de [Wang et al., 2003, Rao et al., 2004]. Ainsi, les ESEs se construisent à partir de la fonction de répartition de la VA considérée, ou plutôt de sa fonction de répartition complémentaire définie par $F_{\mathbf{x}}(\mathbf{u}) = P(\mathbf{x} > \mathbf{u}) = P(\mathbf{x}^{(1)} > \mathbf{u}^{(1)} \text{ et } \mathbf{x}^{(2)} > \mathbf{u}^{(2)})$. Nous notons également $F_{|\mathbf{x}|}(\mathbf{u}) = P(|\mathbf{x}| > \mathbf{u}) = P(|\mathbf{x}^{(1)}| > \mathbf{u}^{(1)} \text{ et } |\mathbf{x}^{(2)}| > \mathbf{u}^{(2)})$. À partir de $F_{|\mathbf{x}|}$, l'ESE d'ordre α s'écrit :

$$\mathcal{M}_{\alpha}(\mathbf{x}) = \left(\oint_{\mathbb{R}^{+2}} F_{|\mathbf{x}|}^{\alpha}(\mathbf{u}) d\mathbf{u} \right)^{\frac{1}{1-\alpha}} \quad (2.47)$$

pour $\alpha \geq 0$, $\alpha \neq 1$. Nous définissons également l'ESEG :

$$\mathcal{S}_{\alpha,\beta}(\mathbf{x}) = \left(\frac{\oint_{\mathbb{R}^{+2}} F_{|\mathbf{x}|}^{\alpha}(\mathbf{u}) d\mathbf{u}}{\oint_{\mathbb{R}^{+2}} F_{|\mathbf{x}|}^{\beta}(\mathbf{u}) d\mathbf{u}} \right)^{\frac{1}{\beta-\alpha}} \quad (2.48)$$

pour $\alpha, \beta \geq 0$ et $\alpha \neq \beta$. Cette dernière expression est obtenue en utilisant la définition de l'entropie exponentielle généralisée [Koski and Persson, 1992].

Il est possible de montrer que $\lim_{\alpha \rightarrow 1} \mathcal{M}_{\alpha}(\mathbf{x})$ coïncide avec l'ECR $\mathcal{E}(\mathbf{x})$ [Wang et al., 2003, Rao et al., 2004] pondérée par $\oint_{\mathbb{R}^{+2}} F_{|\mathbf{x}|}(\mathbf{u}) d\mathbf{u}$. Cette ECR est ainsi donnée par :

$$\mathcal{E}(\mathbf{x}) = - \oint_{\mathbb{R}^{+2}} F_{|\mathbf{x}|}(\mathbf{u}) \log(F_{|\mathbf{x}|}(\mathbf{u})) d\mathbf{u} \quad (2.49)$$

À partir de ces 3 mesures d'entropie, il est possible de dériver des expressions de l'IM en suivant la définition (2.8) qui utilise l'entropie relative plutôt que l'entropie conjointe. Pour cela, les auteurs ont présenté les entropies relatives obtenues à partir des expressions précédentes. Pour les expressions (2.47) et (2.48), ces entropies relatives sont respectivement données par [Liao and Chung, 2006] :

$$\mathcal{M}_{\alpha}(\mathbf{x}|\mathbf{y}) = \left(\oint_{\mathbb{R}^{+2}} F_{\mathbf{x}|\mathbf{y}}^{\alpha}(\mathbf{u}) d\mathbf{u} \right)^{\frac{1}{1-\alpha}} \quad (2.50)$$

pour $\alpha \geq 0$, $\alpha \neq 1$ et

$$\mathcal{S}_{\alpha,\beta}(\mathbf{x}|\mathbf{y}) = \left(\frac{\oint_{\mathbb{R}^{+2}} F_{\mathbf{x}|\mathbf{y}}^{\alpha}(\mathbf{u}) d\mathbf{u}}{\oint_{\mathbb{R}^{+2}} F_{\mathbf{x}|\mathbf{y}}^{\beta}(\mathbf{u}) d\mathbf{u}} \right)^{\frac{1}{\beta-\alpha}} \quad (2.51)$$

pour $\alpha, \beta \geq 0$ et $\alpha \neq \beta$. Pour l'expression de l'ECR (2.49), elle s'écrit :

$$\mathcal{E}(\mathbf{x}|\mathbf{y}) = - \oint_{\mathbb{R}^{+2}} F_{\mathbf{x}|\mathbf{y}}(\mathbf{u}) \log(F_{\mathbf{x}|\mathbf{y}}(\mathbf{u})) d\mathbf{u} \quad (2.52)$$

Dans ces 3 expressions, $F_{\mathbf{x}|\mathbf{y}}(\mathbf{u}) = P(|\mathbf{x}| > \mathbf{x}|\mathbf{y})$. A partir de (2.49) et (2.52) a été définie l'**entropie cumulative résiduelle croisée** (ECRC) [Wang et al., 2003], autrement appelée **IM cumulative** [Drissi et al., 2008] :

$$ECRC(\mathbf{x}|\mathbf{y}) = \mathcal{E}(\mathbf{x}) - E[\mathcal{E}(\mathbf{x}|\mathbf{y})] \quad (2.53)$$

où $E[\mathcal{E}(\mathbf{x}|\mathbf{y})]$ représente l'espérance de (2.52). Pour l'instant, nous avons défini ces mesures avec des vecteurs aléatoires \mathbf{x} et \mathbf{y} à valeurs dans \mathbb{R}^2 . Dans notre contexte de recalage, nous utilisons des VAs $x^{(1)}$ et $x^{(2)}$ monodimensionnelles associées aux images de référence et flottante respectivement. L'expression (2.53) dans notre contexte se réécrit donc :

$$ECRC(x^{(1)}|x^{(2)}) = \mathcal{E}(x^{(1)}) - E[\mathcal{E}(x^{(1)}|x^{(2)})] \quad (2.54)$$

Une différence majeure entre l'IM classique et l'ECRC est que la première est symétrique, contrairement à la seconde. C'est pourquoi une version symétrisée, nommée **entropie cumulative résiduelle croisée symétrique** (ECRCS) dans la littérature, et présentée comme une IM basée sur l'entropie cumulative résiduelle, est préférée dans [Cahill, 2010]. Pour montrer explicitement que cette mesure est une IM, nous la notons IM_{ECRCS} :

$$\begin{aligned} IM_{ECRCS}(x^{(1)}, x^{(2)}) &= \frac{1}{2} \left(ECRC(x^{(1)}|x^{(2)}) + ECRC(x^{(2)}|x^{(1)}) \right) \\ &= \frac{1}{2} \left(\mathcal{E}(x^{(1)}) + \mathcal{E}(x^{(2)}) - E[\mathcal{E}(x^{(1)}|x^{(2)})] - E[\mathcal{E}(x^{(2)}|x^{(1)})] \right) \end{aligned} \quad (2.55)$$

Pour l'ESE et l'ESEG, les formules de l'IM non-symétrique sont apparues dans [Liao and Chung, 2006]. Leurs versions symétrisées sont quant à elles données dans [Cahill, 2010]. Nous les notons respectivement :

$$IM_{ESE}(x^{(1)}, x^{(2)}) = \frac{1}{2} \left(\mathcal{M}_\alpha(x^{(1)}) + \mathcal{M}_\alpha(x^{(2)}) - E[\mathcal{M}_\alpha(x^{(1)}, x^{(2)})] - E[\mathcal{M}_\alpha(x^{(2)}, x^{(1)})] \right) \quad (2.56)$$

et :

$$IM_{ESEG}(x^{(1)}, x^{(2)}) = \frac{1}{2} \left(\mathcal{S}_{\alpha,\beta}(x^{(1)}) + \mathcal{S}_{\alpha,\beta}(x^{(2)}) - E[\mathcal{S}_{\alpha,\beta}(x^{(1)}, x^{(2)})] - E[\mathcal{S}_{\alpha,\beta}(x^{(2)}, x^{(1)})] \right) \quad (2.57)$$

2.6 Discussion au sujet des différentes mesures

Comparer les différentes mesures présentées est une tâche relativement ardue. Tout d'abord, il n'existe pas d'étude comparative de toutes ces mesures. De plus, la mesure de similarité n'est qu'une des composantes d'une méthode de recalage. L'interpolation, le modèle de transformation mais aussi et surtout la méthode d'optimisation choisie vont grandement influencer le résultat final. Dans les différents articles présentant ces mesures, les auteurs ont fait des choix différents au regard de ces composantes. Aussi, comparer des mesures de similarité sans un schéma d'optimisation commun ne présente pas à notre sens un grand intérêt. De plus, les tests menés ne sont pas forcément similaires d'un article à l'autre, ce qui rend encore plus difficile une comparaison à partir des seuls articles originaux.

Nous pouvons néanmoins tirer quelques conclusions de tous les articles présentés dans ce chapitre.

Tout d'abord, presque toutes les nouvelles mesures proposées sont comparées à l'IMN, qui reste encore la référence en matière de mesure de similarité pour le recalage multimodal d'images médicales. Cependant, comme nous l'avons déjà évoqué, peu d'auteurs donnent explicitement l'estimateur de l'IMN utilisé. Dans nos propres expérimentations qui seront présentées au chapitre 4, nous utiliserons également l'IMN comme méthode de référence, mais également l'IM traditionnelle, et nous expliciterons à chaque fois l'estimateur de référence choisi pour la construction de la mesure.

Nous avons trouvé peu d'études comparatives des mesures présentées dans ce chapitre [Pluim et al., 2004, Škerl et al., 2007, Cahill, 2010]. La grande étude sur le recalage d'images médicales multimodales reste celle de [West et al., 1997], qui montre les bonnes performances de l'IMN par rapport à 16 mesures qui sont très différentes les unes des autres. Pour l'évaluation de ces méthodes, une base de données test d'images IRM, TEP, et Scanner a été utilisée. Cette base de données, RIRE (présentée en détail dans l'annexe E), demeure une base de référence pour l'évaluation d'une nouvelle mesure de similarité. Elle présente l'avantage de donner un « gold standard » pour le recalage rigide multimodal en imagerie cérébrale, et il est possible de soumettre les résultats de ses simulations en ligne. Aussi, beaucoup d'auteurs [Studholme et al., 1999, Pluim et al., 2001, Bardera et al., 2006, Liao and Chung, 2006, Cahill et al., 2009] l'utilisent pour valider leur nouvelle mesure de similarité. C'est également cette base de données qui est utilisée dans les articles de comparaison [Pluim et al., 2004, Škerl et al., 2007, Cahill, 2010].

Le premier de ces articles [Pluim et al., 2004] se propose de comparer les f -mesures (2.45) dont fait notamment partie l' IM_α (2.44) en faisant varier le paramètre α . Les résultats montrent que dans certains cas, les f -mesures sont plus efficaces que l'IM classique, mais les auteurs n'ont pas réussi à trouver un paramètre α optimal. L'utilisateur de ces mesures doit donc fixer ce paramètre en fonction des images à recalcr.

L'étude de [Škerl et al., 2007] est sans doute la plus exhaustive, puisqu'elle compare les mesures « classiques » ([Viola and Wells, 1997] (2.20), l'IMN [Studholme et al., 1999] (2.30), le RCE [Maes et al., 1997] (2.31), l'entropie conjointe [Collignon et al., 1995b], [Studholme and Hawkes, 1995] (2.7), le RC [Roche et al., 1998] (1.12), le critère de Woods [Woods et al., 1993] (1.10)) aux mesures basées sur d'autres entropies (l' IM_α [Pluim et al., 2001], et l'IM basé sur l'entropie de Havrda-Charvat [Wachowiak et al., 2003]) ainsi qu'à leur propre mesure, l'IM Multi-Caractéristiques Symétrique (IMMCS), que nous n'avons pas présentée dans ce chapitre mais que le lecteur intéressé pourra trouver dans [Tomazevic et al., 2006]. L'intérêt de cet article réside dans la comparaison de toutes les mesures originelles et des mesures basées sur d'autres entropies sur une base commune. Les résultats ne montrent pas de supériorité très significative d'une mesure par rapport à une autre.

La dernière étude est à mettre à l'actif de [Cahill, 2010] qui compare les différentes mesures d'entropie présentées dans la section 2.5.4, encore une fois sur la base RIRE. Le lecteur intéressé par ces différentes mesures pourra s'y référer, même si les résultats ne montrent pas la supériorité majeure d'une méthode en particulier, l'IM classique donnant toujours de très bons résultats en comparaison des autres mesures.

2.7 Conclusion

De par sa conception, l'IM est un critère de choix pour les applications de recalage multimodal : elle ne présuppose qu'une dépendance statistique entre les images à recalcr. Mais nous

avons aussi vu que la mise en oeuvre de l'estimation de l'information mutuelle peut s'avérer délicate. Il faut en effet faire un choix quant à l'estimation de l'histogramme conjoint, choix qui peut être compliqué puisque des paramètres doivent être fixés, de façon plus ou moins arbitraire. De plus, les estimateurs actuels demandent un nombre important d'échantillons, ce qui les rend coûteux en terme de temps de calcul. Un autre inconvénient majeur est que l'IM n'intègre pas d'information spatiale.

Pour contrer ces inconvénients, des recherches actives sont menées depuis plus de 10 ans. Il nous a semblé intéressant, à travers ce chapitre, de présenter les travaux conduits pour améliorer l'estimation de l'IM. Cela permet tout d'abord de situer l'apport de notre travail, mais surtout de comprendre les nouveaux enjeux du recalage basé sur l'IM. Ainsi, les nouvelles mesures qui se sont développées s'orientent suivant 3 grands axes (même si cette classification n'est pas figée) et cherchent à :

- ajouter de l'information sur les images à recalcr, que ce soit une information spatiale ou statistique ;
- mettre en oeuvre des méthodes d'estimation de l'entropie se substituant à l'histogramme conjoint traditionnellement utilisé ;
- enfin, explorer de nouvelles mesures d'entropie pouvant satisfaire à l'IM.

Les méthodes que nous avons présentées dans ce chapitre s'attachent en général à répondre à un, voire deux des axes de recherche précités.

A travers ce travail de thèse, nous avons également voulu contribuer à ces recherches, en nous appuyant sur ces axes prometteurs. Nous voulions disposer d'une mesure de similarité qui soit à la fois simple à mettre en oeuvre, mais également efficace pour le recalage d'images multimodales.

Pour cela, nous avons suivi une voie nouvelle en traitement d'images. Au départ, nous nous sommes intéressés aux statistiques d'ordre supérieur que sont les cumulants. Ce sont des outils statistiques utilisés classiquement en traitement du signal, qui sont très informatifs. Nous avons donc pensé que ces outils permettraient d'apporter une information statistique supplémentaire sur les images, ce qui pourrait s'inscrire dans le premier axe de recherche que nous avons évoqué.

Puis, à l'aide de ces cumulants, nous avons choisi de développer une approximation de l'IM. Pour cela, nous avons utilisé le développement d'Edgeworth qui permet d'approximer une densité de probabilité à partir de ses cumulants. De cette approximation, il est possible de construire une approximation de l'IM qui s'affranchit de la construction d'histogrammes conjoints. Qui plus est, cette approximation de l'IM ne nécessite aucun paramètre à fixer.

Dans le chapitre 3, nous commencerons par exposer les outils mathématiques nécessaires à la compréhension de ce travail. Puis nous expliquerons le cheminement qui nous a mené de l'utilisation des cumulants à la proposition d'une nouvelle approximation de l'IM. Nous verrons notamment que les mesures utilisées en traitement du signal ne peuvent pas être transposées directement à notre problème de recalage, et que nous y avons donc apporté des modifications substantielles.

Bibliographie

- [Bardera et al., 2006] Bardera, A., Feixas, M., Boada, I., and Sbert, M. (2006). High dimensional normalized mutual information for image registration using random lines. In *International Workshop on Medical Image Registration*.
- [Biesdorf et al., 2009] Biesdorf, A., Worz, S., Kaiser, H.-J., Stippich, C., and Rohr, K. (2009). Hybrid spline-based multimodal registration using local measures for joint entropy and mu-

- tual information. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, volume 5761 of *Lecture Notes in Computer Science*, pages 607–615. Springer Berlin / Heidelberg.
- [Cahill et al., 2008] Cahill, N., Schnabel, J., Noble, J., and Hawkes, D. (2008). Revisiting overlap invariance in medical image alignment. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 1 :1–8.
- [Cahill et al., 2009] Cahill, N., Schnabel, J., Noble, J., and Hawkes, D. (2009). Overlap invariance of cumulative residual entropy measures for multimodal image alignment. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 7259.
- [Cahill, 2010] Cahill, N. D. (2010). Normalized measures of mutual information with general definitions of entropy for multimodal image registration. In *Proceedings of the 4th international conference on Biomedical image registration, WBIR'10*, pages 258–268, Berlin, Heidelberg. Springer-Verlag.
- [Cahill et al., 2006] Cahill, N. D., Williams, C. M., Chen, S., Ray, L. A., and Goodname, M. M. (2006). Incorporating spatial information into entropy estimates to improve multimodal image registration. In *IEEE International Symposium on Biomedical Imaging (ISBI)*, pages 832–835.
- [Collignon et al., 1995a] Collignon, A., Maes, F., Delaere, D., Vandermeulen, D., Suetens, P., and Marchal, G. (1995a). Automated Multi-modality Image Registration based on Information Theory. In *Conference on Information Processing in Medical Imaging*, pages 263–274.
- [Collignon et al., 1995b] Collignon, A., Vandermeulen, D., Suetens, P., and Marchal, G. (1995b). 3D Multi-Modality Medical Image Registration Using Feature Space Clustering. *Lecture Notes in Computer Science*, 905 :195–204.
- [De Nigris et al., 2010] De Nigris, D., Mercier, L., Del Maestro, R., Louis Collins, D., and Arbel, T. (2010). Hierarchical multimodal image registration based on adaptive local mutual information. 6362 :643–651.
- [Drissi et al., 2008] Drissi, N., Chonavel, T., and Boucher, J. M. (2008). Generalized cumulative residual entropy for distributions with unrestricted supports. *Research letters in signal processing*, 2008 :111–115.
- [Hartley, 1928] Hartley, R. (1928). Transmission of information. *Bell System Technical Journal*, 1 :535–563.
- [Hero et al., 2002] Hero, A., Ma, B. Michel, O., and Gorman, J. (2002). Applications of entropic spanning graphs. *IEEE Signal Processing Magazine*, 19(5) :85–95.
- [Hero and Michel, 1999] Hero, A. O. I. and Michel, O. (1999). Asymptotic theory of greedy approximations to minimal k-point random graphs. *IEEE Transactions on Information Theory*, 45 :1921–1938.
- [Holden et al., 2004] Holden, M., Griffin, L. D., Saeed, N., and Hill, D. L. G. (2004). Multi-channel mutual information using scale space. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 797–804.
- [Karaçali, 2007] Karaçali, B. (2007). Information theoretic deformable registration using local image information. *International Journal of Computer Vision*, 72 :219–237.
- [Koski and Persson, 1992] Koski, T. and Persson, L.-E. (1992). Some properties of generalized exponential entropies with applications to data compression. *Information Sciences*, 62 :103–132.
- [Kybic, 2007] Kybic, J. (2007). High-dimensional entropy estimation for finite accuracy data : R-nn entropy estimator. 4584 :569–580.

- [Liao and Chung, 2006] Liao, S. and Chung, A. (2006). Multi-modal image registration using the generalized survival exponential entropy. 4191 :964–971.
- [Loeckx et al., 2009] Loeckx, D., Slagmolen, P., Maes, F., Vandermeulen, D., and Suetens, P. (2009). Nonrigid image registration using conditional mutual information. *Medical Imaging, IEEE Transactions on*, 29(1) :19–29.
- [Ma et al., 2000] Ma, B., Hero, A., Gomtan, J., and Michel, O. (2000). Image registration with minimum spanning tree algorithm. In *International Conference on Image Processing*.
- [Maes et al., 1997] Maes, F., Collignon, A., Vandermeulen, D., Marchal, G., and Suetens, P. (1997). Multimodality image registration by maximization of mutual information. *IEEE Transactions on Medical Imaging*, 16(2) :187–198.
- [Mattes et al., 2003] Mattes, D., Haynor, D. R., Vesselle, H., Lewellen, T. K., and Eubank, W. (2003). PET-CT image registration in the chest using free-form deformations. *IEEE Transactions on Medical Imaging*, 22(1) :120–128.
- [Meyer et al., 1997] Meyer, C. R., Boes, J. L., Kim, B., Bland, P. H., Zasadny, K. R., Kison, P. V., Koral, K., Frey, K. A., and Wahl, R. L. (1997). Demonstration of accuracy and clinical versatility of mutual information for automatic multimodality image fusion using affine and thin-plate spline warped geometric deformations. *Medical Image Analysis*, 1(3) :195 – 206.
- [Moddemeijer, 1999] Moddemeijer, R. (1999). A statistic to estimate the variance of the histogram-based mutual information estimator based on dependent pairs of observations. *Signal Processing*, 75(1) :51–63.
- [Neemuchwala and Hero, 2005] Neemuchwala, H. F. and Hero, A. (2005). *Entropic Graphs for Image Registration*, chapter 6, pages 185–235. Signal Processing and Communications.
- [Parzen, 1962] Parzen, E. (1962). On estimation of a probability density function and mode. *The Annals of Mathematical Statistics*, 33(3) :1065–1076.
- [Pluim et al., 2000] Pluim, J., Maintz, J., and Viergever, M. (2000). Image registration by maximization of combined mutual information and gradient information. *IEEE Transactions on Medical Imaging*, 19(8) :809–814.
- [Pluim et al., 2001] Pluim, J., Maintz, J., and Viergever, M. (2001). f-Information measures in medical image registration. In Sonka, M. and Hanson, K., editors, *SPIE Medical Imaging*, volume 4322, pages 579–587. SPIE press.
- [Pluim et al., 2003] Pluim, J., Maintz, J., and Viergever, M. (2003). Mutual information based registration of medical images : a survey. *IEEE Transactions on Medical Imaging*, 22(8) :986–1004.
- [Pluim et al., 2004] Pluim, J. P. W., Maintz, J. B. A., and Viergever, M. A. (2004). f-information measures in medical image registration. *IEEE Transactions on Medical Imaging*, 23(12) :1508–1516.
- [Rao et al., 2004] Rao, M., Chen, Y., Vemuri, B. C., and Wang, F. (2004). Cumulative residual entropy, a new measure of information. *IEEE Transactions on Information Theory*, 50 :1220–1228.
- [Renyi, 1960] Renyi, A. (1960). On measures of entropy and information. In *Proceedings of the 4th Berkeley Symposium on Mathematics, Statistics and Probability*, pages 547–561.
- [Roche et al., 1998] Roche, A., Malandain, G., Pennec, X., and Ayache, N. (1998). The Correlation Ratio as a New Similarity Measure for Multimodal Image Registration. In W.M. Wells, A. C. and Delp, S., editors, *Medical Image Computing and Computer-Assisted Intervention : MICCAI 1998*, volume 1496, pages 1115–1124. Incs.

- [Rohlfing et al., 2004] Rohlfing, T., Maurer Jr, C., O'Dell, W., and Zhong, J. (2004). Modeling liver motion and deformation during the respiratory cycle using intensity-based nonrigid registration of gated MR images. *Medical physics*, 31(3) :427–432.
- [Rosenblatt, 1956] Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. 27 :832–837.
- [Rueckert et al., 2000] Rueckert, D., Clarkson, M. J., Hill, D. L. G., and Hawkes, D. J. (2000). Non-rigid registration using higher-order mutual information. *SPIE proceedings series*, 3979(2) :438–447.
- [Rueckert et al., 1999] Rueckert, D., Sonoda, L., Hayes, C., Hill, D., Leach, M., and Hawkes, D. (1999). Nonrigid registration using free-form deformations : Application to breast MR images. *IEEE Transactions on Medical Imaging*, 18(8) :712–721.
- [Russakoff et al., 2004] Russakoff, D. B., Tomasi, C., Rohlfing, T., Maurer, C. R., and Jr. (2004). Image similarity using mutual information of regions. In *8th European Conference on Computer Vision (ECCV)*, pages 596–607. Springer.
- [Sabuncu, 2006] Sabuncu, M. R. (2006). *Entropy-based Image Registration*. PhD thesis, Princeton University.
- [Sabuncu and Ramadge, 2003] Sabuncu, M. R. and Ramadge, P. J. (2003). Spatial information in entropy-based image registration. In *Proceedings of 2nd Workshop in Biomedical Image Registration (WBIR03), LNCS 2717. (2003) 132-141*.
- [Sabuncu and Ramadge, 2008] Sabuncu, M. R. and Ramadge, P. J. (2008). Using spanning graphs for efficient image registration. *IEEE Transactions on Image Processing*, 17 :788–797.
- [Sarrut, 2000] Sarrut, D. (2000). *Recalage multimodal et plate-forme d'imagerie médicale à accès distant*. PhD thesis, Université Lumière, Lyon, France.
- [Shannon, 1948] Shannon, C. (1948). A mathematical theory of communication. *Bell System Technical Journal*, 27 :379–423/623–656.
- [Shannon and Weaver, 1949] Shannon, C. E. and Weaver, W. (1949). *The Mathematical Theory of Communication*. University of Illinois Press.
- [Slomka et al., 2003] Slomka, P., Dey, D., Przetak, C., Aladl, U., and Baum, R. (2003). Automated 3-dimensional registration of stand-alone 18F-FDG whole-body PET with CT. *Journal of Nuclear Medicine*, 44(7) :1156–67.
- [Staring et al., 2009] Staring, M., van der Heide, U., Klein, S., Viergever, M., and Pluim, J. (2009). Registration of cervical MRI using multifeature mutual information. *IEEE Transactions on Medical Imaging*, 28(9) :1412–1421.
- [Studholme et al., 2006] Studholme, C., Drapaca, C., Iordanova, B., and Cardenas, V. (2006). Deformation-based mapping of volume change from serial brain MRI in the presence of local tissue contrast change. *IEEE Transactions on Medical Imaging*, 25(5) :626–639.
- [Studholme et al., 1997] Studholme, C., Hill, D. L. G., and Hawkes, D. J. (1997). Automated three-dimensional registration of magnetic resonance and positron emission tomography brain images by multiresolution optimization of voxel similarity measures. *Medical Physics*, 24(1) :25–35.
- [Studholme et al., 1999] Studholme, C., Hill, D. L. G., and Hawkes, D. J. (1999). An overlap invariant entropy measure of 3D medical image alignment. *Pattern Recognition*, 32(1) :71–86.
- [Studholme and Hawkes, 1995] Studholme, C. Hill, D. L. G. and Hawkes, D. J. (1995). Multiresolution voxel similarity measures for MR-PET registration. In Bizais, Y. and Barillot,

- C., editors, *Information processing in medical imaging*, Computational imaging and vision, pages 287–298.
- [Thévenaz and Unser, 1997] Thévenaz, P. and Unser, M. (1997). Spline pyramids for inter-modal image registration using mutual information. In *Proceedings of the SPIE Conference on Mathematical Imaging : Wavelet Applications in Signal and Image Processing V*, volume 3169, pages 236–247, San Diego CA, USA.
- [Thevenaz and Unser, 2000] Thevenaz, P. and Unser, M. (2000). Optimization of mutual information for multiresolution image registration. *IEEE Transactions on Image Processing*, 9 :2083–2099.
- [Tomazevic et al., 2006] Tomazevic, D., Likar, B., and Pernus, F. (2006). 3-D/2-D registration by integrating 2-D information in 3-D. *Medical Imaging, IEEE Transactions on*, 25(1) :17–27.
- [Unser et al., 1991] Unser, M., Aldroubi, A., and Eden, M. (1991). Fast b-spline transforms for continuous image representation and interpolation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13 :277–285.
- [Unser et al., 1993a] Unser, M., Aldroubi, A., and Eden, M. (1993a). B-Spline signal processing : Part I-Theory. *IEEE Transactions on Signal Processing*, 41(2) :821–833.
- [Unser et al., 1993b] Unser, M., Aldroubi, A., and Eden, M. (1993b). B-Spline signal processing : Part II-Efficient design and applications. *IEEE Transactions on Signal Processing*, 41(2) :834–848.
- [Vajda, 1989] Vajda, I. (1989). *Theory of statistical inference and information*. Springer.
- [Viola, 1995] Viola, P. (1995). *Alignement by Maximization of Mutual Information*. PhD thesis, Massachusetts Institute of Technology Artificial Intelligence Laboratory, Boston, USA.
- [Viola and Wells, 1997] Viola, P. and Wells, W. (1997). Alignement by maximization of Mutual Information. *International Journal of Computer Vision*, 24(2) :137–154.
- [Škerl et al., 2007] Škerl, D., Likar, B., Fitzpatrick, J. M., and Pernuš, F. (2007). Comparative evaluation of similarity measures for the rigid registration of multi-modal head images. *Physics in Medicine and Biology*, 52(18) :5587.
- [Wachowiak et al., 2003] Wachowiak, M. P., Smolikova, R., Tourassi, G. D., and Elmaghraby, A. S. (2003). Similarity metrics based on nonadditive entropies for 2D-3D multimodal biomedical image registration. In *SPIE Medical Imaging*.
- [Wang et al., 2003] Wang, F., Vemuri, B. C., Rao, M., and Chen, Y. (2003). Cumulative residual entropy, a new measure of information & its application to image alignment. In *Proceedings of the Ninth IEEE International Conference on Computer Vision - Volume 2*, ICCV '03, pages 548–, Washington, DC, USA. IEEE Computer Society.
- [West et al., 1997] West, J., Fitzpatrick, J. M., Wang, M. Y., Dawant, B. M., Maurer, C. R., Kessler, R. M., Maciunas, R. J., Barillot, C., Lemoine, D., Collignon, A., Maes, F., Sumanaweera, T. S., Harkness, B., Hemler, P. F., Hill, D. L. G., Hawkes, D. J., Studholme, C., Maintz, J. B. A., Viergever, M. A., Mal, G., Pennec, X., Noz, M. E., Maguire, G. Q., Pollock, M., Pelizzari, C. A., Robb, R. A., Hanson, D., and Woods, R. P. (1997). Comparison and Evaluation of Retrospective Intermodality Image Registration Techniques. *Journal of Computer Assisted Tomography*, 21(4) :554–566.
- [Woods et al., 1993] Woods, R., Mazziota, J., and Cherry, S. (1993). MRI-PET Registration with Automated Algorithm. *Journal of Computer Assisted Tomography*, 17(4) :536–546.
- [Zografos and Nadarajah, 2005] Zografos, K. and Nadarajah, S. (2005). Survival exponential entropies. *IEEE Transactions on Information Theory*, 51 :1239–1246.

Chapitre 3

Approximations de l'Information Mutuelle basées sur les cumulants

Au cours des chapitres précédents, nous avons vu que l'IM est une mesure de choix en recalage d'images médicales multimodales. Elle tient notamment son efficacité du fait qu'elle permet de travailler directement avec les intensités des deux images à recaler, sans nécessiter l'utilisation de caractéristiques particulières des images. Nous avons vu également que cette mesure de similarité posait des problèmes d'un point de vue pratique, puisque son estimation n'est pas toujours aisée, et peut notamment induire des biais d'estimation. La principale faiblesse de cette mesure est sans doute la construction de l'histogramme conjoint nécessaire à son estimation. Le chapitre précédent décrivait les avancées et les modifications substantielles apportées à l'IM depuis les premiers travaux utilisant cette mesure en recalage d'images [Collignon et al., 1995, Viola, 1995]. Dans ce chapitre, qui constitue la contribution principale de cette thèse, nous présentons de nouvelles mesures de similarité basées sur les cumulants, dont certaines d'entre-elles cherchent à approximer l'IM. Dans un premier temps, nous donnerons les outils mathématiques nécessaires à la compréhension de ce chapitre. Puis, nous justifierons le choix d'utiliser les cumulants pour construire une mesure de similarité en recalage d'images médicales. Nous montrerons notamment que ces cumulants sont utilisés en traitement du signal pour construire une approximation de l'IM basée sur un développement d'Edgeworth. Nous verrons que ces approximations ne peuvent être utilisées dans notre contexte applicatif, celui-ci étant différent des problèmes rencontrés en traitement de signal. C'est pourquoi nous présenterons une nouvelle approximation de l'IM basée sur ce même développement d'Edgeworth, mais compatible cette fois avec notre problème de recalage.

3.1 Prérequis mathématiques

Avant de rentrer dans les développements mathématiques que nous avons conduits, il faut se doter des outils nécessaires à leur compréhension. Nous nous intéresserons tout d'abord aux **polynômes et tenseurs d'Hermite**, dont nous exploiterons les propriétés d'orthogonalité. Les **cumulants**, qui sont au centre de ce chapitre, auront ensuite toute notre attention. Nous verrons également le lien qui les unissent aux **moments**. Dans toute cette partie, nous respecterons le formalisme de [McCullagh, 1987, Barndorff-Nielsen and Cox, 1989], qui présente l'avantage de proposer des relations simples lorsque l'on étend les formules au cas N -dimensionnel, ce qui sera utile dans notre démonstration. Finalement, nous considérons souvent dans nos développements que nos variables sont centrées-réduites, et nous le précisons le cas échéant. En pratique, cela consiste à centrer et normaliser les valeurs en niveau de

gris des pixels (ou voxels) des images. C'est un pré-traitement classiquement utilisé en recalage d'images utilisant des mesures statistiques. Ceci offre l'avantage de simplifier l'expression de ces dernières.

3.1.1 Polynômes et tenseurs d'Hermite

Les polynômes d'Hermite ont été nommés ainsi en l'honneur du mathématicien français Charles Hermite. Ces polynômes font partie de la famille des polynômes orthogonaux, et leurs propriétés d'orthogonalité vont nous être utiles par la suite. Ici, nous nous intéresserons à quelques idées importantes concernant ces polynômes, dans le cas mono et multidimensionnel, mais tous les résultats utilisés dans cette thèse sont présentés en détail dans l'annexe D. Dans le cas multivarié, les polynômes seront appelés tenseurs d'Hermite. Dans ce qui suit, nous nous intéressons aux polynômes d'Hermite associés à la distribution normale. Les notations ainsi que la majorité des définitions utilisées sont issues de [McCullagh, 1987] et [Barndorff-Nielsen and Cox, 1989].

3.1.1.1 Polynômes d'Hermite

Pour construire les polynômes d'Hermite, nous utilisons une distribution normale, ou gaussienne, monovariée d'espérance $\mathbb{E}[x]$ et de variance $\text{Var}(x)$. Cette distribution ϕ_x est définie sur \mathbb{R} par :

$$\phi_x(u) = \frac{1}{\sqrt{2\pi\text{Var}(x)}} e^{-(u-\mathbb{E}[x])^2/(2\text{Var}(x))} \quad (3.1)$$

La loi normale centrée-réduite correspondante, qui nous servira par la suite, est quant à elle définie par :

$$\phi_x(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} \quad (3.2)$$

Les polynômes d'Hermite associés à cette distribution sont définis par l'identité [Barndorff-Nielsen and Cox, 1989] :

$$\phi_x(u)H_k(u) = (-1)^k \frac{d^k \phi_x(u)}{du^k} \quad (3.3)$$

On voit immédiatement que cette identité définit un polynôme de degré k de terme dominant 1. L'orthogonalité peut être démontrée par intégration par parties. Les premiers polynômes d'Hermite univariés sont donnés par :

$$\begin{aligned} H_0(u) &= 1 \\ H_1(u) &= u \\ H_2(u) &= u^2 - 1 \\ H_3(u) &= u^3 - 3u \\ H_4(u) &= u^4 - 6u^2 + 3 \\ H_5(u) &= u^5 - 10u^3 + 15u \\ H_6(u) &= u^6 - 15u^4 + 45u^2 - 15 \end{aligned} \quad (3.4)$$

Ces polynômes ont la propriété intéressante de former un système orthonormal dans le sens où ([Abramowitz and Stegun, 1964], p.775) :

$$\int_{\mathbb{R}} \phi_x(u) H_i(u) H_j(u) du = \begin{cases} i! & \text{si } i = j \\ 0 & \text{autrement} \end{cases} \quad (3.5)$$

3.1.1.2 Tenseurs d'Hermite

Précédemment, nous avons vu que les polynômes d'Hermite étaient définis à partir de la distribution normale. Les tenseurs d'Hermite ou polynômes d'Hermite multivariés vont quand à eux être associés à une loi normale multidimensionnelle, généralisation multidimensionnelle de la loi normale. Pour notre problème de recalage entre 2 VAs $x^{(1)}$ et $x^{(2)}$, nous n'avons besoin que des polynômes d'Hermite bidimensionnels associés à une distribution normale bivariée. Cette dernière est définie par :

$$\phi_{\mathbf{x}}(\mathbf{u}) = \frac{1}{(2\pi)|\boldsymbol{\kappa}_{\mathbf{x}}|^{1/2}} e^{-\frac{1}{2}(\mathbf{u}-\mathbb{E}[\mathbf{x}])^T \boldsymbol{\kappa}_{\mathbf{x}}^{-1}(\mathbf{u}-\mathbb{E}[\mathbf{x}])} \quad (3.6)$$

où $\boldsymbol{\kappa}_{\mathbf{x}} = \mathbb{E}[\mathbf{x}\mathbf{x}^T] - \mathbb{E}[\mathbf{x}]\mathbb{E}[\mathbf{x}]^T$ est la matrice de covariance de \mathbf{x} dont les éléments notés $\kappa_{\mathbf{x}}^{i,j}$ sont les cumulants d'ordre 2 de \mathbf{x} que nous étudierons en détail dans la section suivante 3.1.2 et où $\boldsymbol{\kappa}_{\mathbf{x}}^{-1}$ est l'inverse de cette matrice de covariance dont les éléments sont notés $\kappa_{i,j}^{\mathbf{x}}$. Nous suivons ici encore une fois les notations de [Barndorff-Nielsen and Cox, 1989] qui utilise les systèmes covariants (indices en bas) et contravariants (indices en haut). De plus, nous donnons dès maintenant la distribution normale bivariée centrée-réduite qui va nous être utile lorsque nous présenterons le développement d'Edgeworth. Elle s'écrit :

$$\phi_{\mathbf{x}}(\mathbf{u}) = \frac{1}{(2\pi)|\boldsymbol{\kappa}_{\mathbf{x}}|^{1/2}} e^{-\frac{1}{2}(\mathbf{u}^T \boldsymbol{\kappa}_{\mathbf{x}}^{-1} \mathbf{u})} \quad (3.7)$$

où les deux matrices $\boldsymbol{\kappa}_{\mathbf{x}}$ et $\boldsymbol{\kappa}_{\mathbf{x}}^{-1}$, de taille (2×2) , et sous l'hypothèse que les données sont centrées-réduites, sont données respectivement par :

$$\boldsymbol{\kappa}_{\mathbf{x}} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \quad (3.8)$$

et :

$$\boldsymbol{\kappa}_{\mathbf{x}}^{-1} = \frac{1}{1-\rho^2} \begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix} \quad (3.9)$$

avec $\rho = \kappa_{\mathbf{x}}^{1,2}$ la covariance entre $x^{(1)}$ and $x^{(2)}$ comprise entre -1 et 1 .

Les polynômes d'Hermite covariants associés à la loi normale bidimensionnelle, notés $H_{i_1 \dots i_k}$, sont définis par :

$$\phi_{\mathbf{x}}(\mathbf{u}) H_{i_1 \dots i_k}(\mathbf{u}) = (-1)^k \partial_{i_1} \dots \partial_{i_k} \phi_{\mathbf{x}}(\mathbf{u}) \quad (3.10)$$

où $\partial_i = \partial/\partial x^{(i)}$. On notera le parallèle direct entre cette définition et celle donnant les polynômes d'Hermite univariés (3.3). Le premier tenseur d'Hermite covariant est donné par :

$$H_{i_1}(\mathbf{u}) = \kappa_{i_1,1}^{\mathbf{x}} u^{(1)} + \kappa_{i_1,2}^{\mathbf{x}} u^{(2)} \quad (3.11)$$

et la relation de récurrence :

$$\partial_i H_{i_1 \dots i_k}(\mathbf{u}) = \kappa_{i,i_1}^{\mathbf{x}} H_{i_2 \dots i_k}(\mathbf{u}) [k] \quad (3.12)$$

permet de retrouver les tenseurs suivants. Nous donnons ici, en posant $\mathbf{u} = (u_1 \dots u_k)^\top$, les tenseurs qui nous seront utiles dans la suite de la démonstration ([Barndorff-Nielsen and Cox, 1989], p.150) :

$$\begin{aligned}
H_{i_1}(\mathbf{u}) &= u_{i_1} \\
H_{i_1 i_2}(\mathbf{u}) &= u_{i_1} u_{i_2} - \kappa_{i_1, i_2}^{\mathbf{x}} \\
H_{i_1 i_2 i_3}(\mathbf{u}) &= u_{i_1} u_{i_2} u_{i_3} - [3] \kappa_{i_1, i_2}^{\mathbf{x}} u_{i_3} \\
H_{i_1 i_2 i_3 i_4}(\mathbf{u}) &= u_{i_1} u_{i_2} u_{i_3} u_{i_4} - [6] \kappa_{i_1, i_2}^{\mathbf{x}} u_{i_3} u_{i_4} + [3] \kappa_{i_1, i_2}^{\mathbf{x}} \kappa_{i_3, i_4}^{\mathbf{x}} \\
H_{i_1 i_2 i_3 i_4 i_5}(\mathbf{u}) &= u_{i_1} u_{i_2} u_{i_3} u_{i_4} u_{i_5} - [10] \kappa_{i_1, i_2}^{\mathbf{x}} u_{i_3} u_{i_4} u_{i_5} + [15] \kappa_{i_1, i_2}^{\mathbf{x}} \kappa_{i_3, i_4}^{\mathbf{x}} u_{i_5} \\
H_{i_1 i_2 i_3 i_4 i_5 i_6}(\mathbf{u}) &= u_{i_1} u_{i_2} u_{i_3} u_{i_4} u_{i_5} u_{i_6} - [15] \kappa_{i_1, i_2}^{\mathbf{x}} u_{i_3} u_{i_4} u_{i_5} u_{i_6} + [45] \kappa_{i_1, i_2}^{\mathbf{x}} \kappa_{i_3, i_4}^{\mathbf{x}} u_{i_5} u_{i_6} \\
&\quad - [15] \kappa_{i_1, i_2}^{\mathbf{x}} \kappa_{i_3, i_4}^{\mathbf{x}} \kappa_{i_5, i_6}^{\mathbf{x}}
\end{aligned} \tag{3.13}$$

où $[d]\{(i_1, \dots, i_\ell)\}$ désigne la notation crochet de McCullagh [McCullagh, 1987]. Le nombre d apparaissant entre crochets nous dit que la quantité $[d]\{(i_1, \dots, i_\ell)\}$ correspond à une somme de d termes se déduisant de $\{(i_1, \dots, i_\ell)\}$ en permutant les indices i_1, \dots, i_ℓ . A titre d'exemple, nous pouvons écrire :

$$[3] \kappa_{i_1, i_2}^{\mathbf{x}} u_{i_3} = \kappa_{i_1, i_2}^{\mathbf{x}} u_{i_3} + \kappa_{i_1, i_3}^{\mathbf{x}} u_{i_2} + \kappa_{i_2, i_3}^{\mathbf{x}} u_{i_1} \tag{3.14}$$

Enfin, comme dans le cas monovarié, la propriété d'intérêt est l'orthogonalité qui s'exprime par [McCullagh, 1987] :

$$\oint_{\mathbb{R}^k} H_{i_1, \dots, i_m}(\mathbf{u}) H_{j_1, \dots, j_\ell}(\mathbf{u}) \phi_{\mathbf{x}}(\mathbf{u}) d\mathbf{u} = \begin{cases} [k!] \kappa_{i_1, j_1}^{\mathbf{x}} \dots \kappa_{i_m, j_m}^{\mathbf{x}} & \text{si } \ell = m \\ 0 & \text{autrement} \end{cases} \tag{3.15}$$

3.1.2 Outils statistiques : Moments et Cumulants

3.1.2.1 Moments et cumulants d'une variable aléatoire réelle

Nous allons rappeler dans cette partie la définition des moments et cumulants d'ordre q ($q \geq 0$). Mais commençons tout d'abord par la définition des fonctions caractéristiques de première et seconde espèces [Lacoume et al., 1997]. Nous nous placerons dans le cas où la fonction de répartition de la variable considérée est dérivable bien que cela ne soit pas une condition nécessaire à l'existence des deux fonctions caractéristiques.

Soit x une VA réelle admettant une densité de probabilité p_x , sa fonction caractéristique de première espèce est définie par :

$$\Phi_x(v) = \mathbb{E}[e^{ivx}] = \int_{\mathbb{R}} e^{ivx} p_x(x) dx \tag{3.16}$$

où $\mathbb{E}[x]$ désigne l'espérance mathématique de x , et $i^2 = -1$.

De ce fait, la fonction Φ_x correspond à la conjuguée complexe de la transformée de Fourier de p_x . Il est donc équivalent de spécifier p_x ou Φ_x pour une VA, puisque ces deux fonctions sont liées de manière univoque par une transformée (inverse) de Fourier. Par ailleurs, il est à noter que la fonction caractéristique de première espèce est continue en tout point et vaut un à l'origine. Elle est donc non nulle dans un voisinage de l'origine, sur lequel on pourra définir son logarithme népérien, nous donnant ainsi la fonction caractéristique de seconde espèce. Celle-ci

est définie par $\Psi_x = \log(\Phi_x)$. On appelle ces fonctions « caractéristiques » car en un sens elles caractérisent la distribution de manière unique.

Dans la suite, nous préférons aux fonctions caractéristiques les fonctions génératrices des moments et des cumulants, qui permettent également de définir ses outils statistiques et qui sont de plus utilisées pour construire le développement d'Edgeworth ([Barndorff-Nielsen and Cox, 1989], p.90). La fonction génératrice des moments s'exprime par :

$$\mathcal{M}_x(v) = \mathbb{E}[e^{vx}] = \int_{\mathbb{R}} e^{vu} p_x(u) du \quad (3.17)$$

lorsque son espérance existe. Elle est utilisée pour générer les moments associées à la distribution de probabilité de la VA x . La valeur $\mathcal{M}_x(v)$, quand v est remplacé par $-v$, est appelée transformée de Laplace pour la distribution de probabilité de x [Kolassa, 1994]. Notons le parallèle direct avec la fonction caractéristique de première espèce (3.16) : $\mathcal{M}_x(iv) = \Phi_x(v)$. Cependant, contrairement à la fonction caractéristique de première espèce, la fonction génératrice des moments n'existe pas toujours. Elle nécessite que $\mathbb{E}[e^{vx}]$ existe, c'est-à-dire que l'intégrale correspondante soit convergente.

La fonction génératrice des cumulants, quant à elle, est donnée par :

$$\mathcal{K}_x(v) = \log(\mathcal{M}_x(v)) \quad (3.18)$$

Les moments et cumulants se déduisent alors des deux fonctions génératrices définies ci-dessus. Le moment d'ordre q d'une VA réelle admettant la fonction génératrice des moments \mathcal{M}_x est la quantité :

$$\mu_{q,x} = \mathbb{E}[x^q] = \left. \frac{d^q \mathcal{M}_x(v)}{dv^q} \right|_{v=0} \quad (3.19)$$

Par la suite, nous noterons le moment d'ordre q par μ_q afin d'alléger les notations, la dépendance à la VA x étant implicite.

De la même manière, une variable aléatoire réelle x admettant la fonction génératrice des cumulants \mathcal{K}_x aura un cumulant d'ordre q défini par :

$$\kappa_{q,x} = \text{Cum}\left\{ \underbrace{x, \dots, x}_{q \text{ termes}} \right\} = \left. \frac{d^q \mathcal{K}_x(v)}{dv^q} \right|_{v=0} \quad (3.20)$$

Comme pour les moments, la dépendance à la VA x sera implicite par la suite et nous noterons donc κ_q le cumulant d'ordre q .

Les relations unissant les cumulants d'ordre q aux moments d'ordre q , pour $1 \leq q \leq 6$ et pour une VA x centrée, sont données par :

$$\begin{aligned} \kappa_1 &= \mu_1 = 0 \\ \kappa_2 &= \mu_2 = \sigma^2 \\ \kappa_3 &= \mu_3 \\ \kappa_4 &= \mu_4 - 3\mu_2^2 \\ \kappa_5 &= \mu_5 - 10\mu_3\mu_2 \\ \kappa_6 &= \mu_6 - 15\mu_4\mu_2 - 10\mu_3^2 + 30\mu_2^3 \end{aligned} \quad (3.21)$$

$$(3.22)$$

Quand les variables sont centrées-réduites, les 6 premiers cumulants s'expriment en fonction des moments par [Stuart and Ord, 2006] :

$$\begin{aligned}\kappa_1 &= \mu_1 = 0 \\ \kappa_2 &= \mu_2 = 1 \\ \kappa_3 &= \mu_3 \\ \kappa_4 &= \mu_4 - 3\end{aligned}\tag{3.23}$$

$$\begin{aligned}\kappa_5 &= \mu_5 - 10\mu_3 \\ \kappa_6 &= \mu_6 - 15\mu_4 - 10\mu_3^2 + 30\end{aligned}\tag{3.24}$$

Les 2 premiers cumulants sont appelés moyenne et variance. De la même manière, nous définissons l'**asymétrie** (ou skewness) et le **kurtosis**, qui s'expriment en fonction des cumulants. L'asymétrie est définie par :

$$\gamma_1 = \frac{\kappa_3}{\kappa_2^{3/2}}\tag{3.25}$$

Dans le cas de variables centrées-réduites, nous avons directement $\gamma_1 = \kappa_3$. L'asymétrie, comme son nom l'indique, va permettre de mesurer le caractère asymétrique d'une distribution. Il ajoute donc par rapport aux moments une information supplémentaire sur la distribution considérée.

Le kurtosis, quant à lui, est défini par :

$$\gamma_2 = \frac{\kappa_4}{\kappa_2^2}\tag{3.26}$$

Encore une fois, quand les variables sont centrées-réduites, nous avons $\gamma_2 = \kappa_4$. Le kurtosis va mesurer l'applatissage d'une distribution. L'introduction des cumulants va donc avoir un rôle important. Cela va permettre de caractériser les distributions considérées notamment au travers d'un développement connu sous le nom d'Edgeworth, comme nous le verrons par la suite.

3.1.2.2 Extension au cas multidimensionnel

Nous pouvons étendre les définitions des fonctions génératrices des moments et des cumulants au cas multivarié, donnant ainsi $\mathcal{M}_{\mathbf{x}}$ et $\mathcal{K}_{\mathbf{x}}$ pour un vecteur aléatoire $\mathbf{x} = (x^{(1)}, \dots, x^{(N)})^\top$ de dimension N . Les moments multivariés $\mu_{\mathbf{x}}^{i_1, \dots, i_N}$ de \mathbf{x} s'écriront alors en fonction de la fonction génératrice des moments multivariée par :

$$\mu_{\mathbf{x}}^{i_1, \dots, i_N} = \mathbb{E}[x^{i_1} \dots x^{i_N}] = \left. \frac{\partial^q \mathcal{M}_{\mathbf{x}}(\mathbf{v})}{\partial v_{i_1}^q \dots \partial v_{i_N}^q} \right|_{\mathbf{v}=\mathbf{0}}\tag{3.27}$$

Les cumulants multidimensionnels s'écrivent quant à eux $\kappa_{\mathbf{x}}^{i_1, \dots, i_N}$ et sont obtenus de la même manière en utilisant cette fois la fonction génératrice des cumulants multivariée $\mathcal{K}_{\mathbf{x}}$. Encore une fois, nous considérons la dépendance au vecteur aléatoire \mathbf{x} implicite dans ce qui suit et noterons donc μ^{i_1, \dots, i_N} les moments et κ^{i_1, \dots, i_N} les cumulants multivariés. Notons la différence dans les notations employées pour les moments et les cumulants par rapport au cas monodimensionnel. Ceci est fait pour respecter le formalisme en vigueur, ainsi que pour simplifier les

notations. Pour faire le lien entre les 2 systèmes de notations, il est à noter que $\mu^1 = \mu_1$ par exemple.

Nous pouvons donner les relations entre les cumulants multidimensionnels de second, troisième et quatrième ordre en fonction des moments multidimensionnels par :

$$\begin{aligned}\kappa^{i_1, i_2} &= \mu^{i_1, i_2} - \mu^{i_1} \mu^{i_2} \\ \kappa^{i_1, i_2, i_3} &= \mu^{i_1, i_2, i_3} - [3] \mu^{i_1} \mu^{i_2, i_3} + 2 \mu^{i_1} \mu^{i_2} \mu^{i_3} \\ \kappa^{i_1, i_2, i_3, i_4} &= \mu^{i_1, i_2, i_3, i_4} - [4] \mu^{i_1} \mu^{i_2, i_3, i_4} - [3] \mu^{i_1, i_2} \mu^{i_3, i_4} + 2[6] \mu^{i_1} \mu^{i_2} \mu^{i_3, i_4} - 6 \mu^{i_1} \mu^{i_2} \mu^{i_3} \mu^{i_4}\end{aligned}\tag{3.28}$$

Dans le cas particulier de variables centrées-réduites, nous aurons $\kappa^1 = \kappa^2 = 0$, $\kappa^{1,1} = \kappa^{2,2} = 1$ et $\kappa^{1,2} = \kappa^{2,1} = \rho$ avec ρ compris entre -1 et 1 . Les 2 dernières relations découlent directement de la matrice de covariance (3.8) de la section précédente. On peut donc réécrire (3.28) comme :

$$\begin{aligned}\kappa^{i_1, i_2} &= \mu^{i_1, i_2} \\ &= \mathbb{E}[x^{(i_1)} x^{(i_2)}] \\ \kappa^{i_1, i_2, i_3} &= \mu^{i_1, i_2, i_3} \\ &= \mathbb{E}[x^{(i_1)} x^{(i_2)} x^{(i_3)}] \\ \kappa^{i_1, i_2, i_3, i_4} &= \mu^{i_1, i_2, i_3, i_4} - [3] \mu^{i_1, i_2} \mu^{i_3, i_4} \\ &= \mathbb{E}[x^{(i_1)} x^{(i_2)} x^{(i_3)} x^{(i_4)}] - [3] \mathbb{E}[x^{(i_1)} x^{(i_2)}] \mathbb{E}[x^{(i_3)} x^{(i_4)}]\end{aligned}\tag{3.29}$$

Enfin, dans ce qui suit, nous n'aurons besoin que de certains cumulants que nous pouvons d'ores et déjà écrire, en fonction de $\mathbf{x} = (x^{(1)}, x^{(2)})^\top$ dont les composantes sont supposées centrées-réduites :

$$\begin{aligned}\kappa^{1,1,1} &= \mathbb{E}[(x^{(1)})^3] \\ \kappa^{1,1,2} &= \mathbb{E}[(x^{(1)})^2 x^{(2)}] \\ \kappa^{1,2,2} &= \mathbb{E}[x^{(1)} (x^{(2)})^2] \\ \kappa^{2,2,2} &= \mathbb{E}[(x^{(2)})^3] \\ \kappa^{1,1,1,1} &= \mathbb{E}[(x^{(1)})^4] - 3 \\ \kappa^{1,1,1,2} &= \mathbb{E}[(x^{(1)})^3 x^{(2)}] - 3\rho \\ \kappa^{1,1,2,2} &= \mathbb{E}[(x^{(1)})^2 (x^{(2)})^2] - 1 - 2\rho^2 \\ \kappa^{1,2,2,2} &= \mathbb{E}[x^{(1)} (x^{(2)})^3] - 3\rho \\ \kappa^{2,2,2,2} &= \mathbb{E}[(x^{(2)})^4] - 3\end{aligned}\tag{3.30}$$

3.1.2.3 Propriétés des moments et cumulants

Les moments et cumulants possèdent un grand nombre de propriétés (voir [McCullagh, 1987, Lacoue et al., 1997] pour plus de détails). Nous ne présentons ici que les deux propriétés qui nous seront utiles par la suite :

Propriété 1 (*Indépendance*) Si $x^{(1)}$ et $x^{(2)}$ sont deux variables aléatoires indépendantes, alors les cumulants croisés d'un groupe de variables aléatoires comprenant $x^{(1)}$ et $x^{(2)}$ sont nuls.

Propriété 2 (*Gaussiannité*) Si x est une variable aléatoire gaussienne, alors tous ses cumulants d'ordre strictement supérieur à deux sont nuls.

3.1.2.4 Estimation des cumulants

En pratique, les cumulants ne peuvent être calculés exactement, et doivent donc être estimés à partir des réalisations des variables aléatoires mises en jeu, c'est à dire dans notre contexte à partir des intensités de certains pixels des images à recaler. En particulier, si nous considérons $\{x^{(i)}(\nu)\}$ un ensemble d'intensités de l'image i représentant des réalisations indépendantes de la VA centrée $x^{(i)}$, nous pouvons écrire :

$$\begin{aligned}\kappa^{i_1, i_2} &\approx \frac{1}{L} \sum_{\ell=1}^L x^{(i_1)}[\ell] x^{(i_2)}[\ell] \\ \kappa^{i_1, i_2, i_3} &\approx \frac{1}{L} \sum_{\ell=1}^L x^{(i_1)}[\ell] x^{(i_2)}[\ell] x^{(i_3)}[\ell] \\ \kappa^{i_1, i_2, i_3, i_4} &\approx \frac{1}{L} \sum_{\ell=1}^L x^{(i_1)}[\ell] x^{(i_2)}[\ell] x^{(i_3)}[\ell] x^{(i_4)}[\ell] \\ &\quad - [3] \left(\frac{1}{L} \sum_{\ell=1}^L x^{(i_1)}[\ell] x^{(i_2)}[\ell] \right) \left(\frac{1}{L} \sum_{\ell=1}^L x^{(i_3)}[\ell] x^{(i_4)}[\ell] \right)\end{aligned}\tag{3.31}$$

Ceci conduit à faire l'hypothèse d'indépendance des réalisations des VAs qui rentrent en jeu dans le calcul, c'est à dire à considérer que chaque intensité d'un pixel dans une image est indépendante des intensités des pixels voisins. Nous avons conscience qu'une telle hypothèse est assez forte et sans doute globalement fautive, puisque les images portent une information locale, comme nous l'avons vu dans le chapitre précédent 2. Par conséquent, des pixels voisins ne sont pas nécessairement indépendants. Cependant en pratique, nous ne considérons pour le calcul des cumulants qu'un sous-ensemble très restreint des pixels des images, comme nous l'étudierons en détail au chapitre 4, sous-ensemble tiré aléatoirement de manière uniforme dans l'ensemble total des pixels disponibles. Ceci tend à garantir une probabilité petite de choisir des pixels voisins.

3.2 Cumulants en traitement du signal

Les cumulants ont été beaucoup utilisés en traitement du signal car ils offrent la possibilité de caractériser finement des distributions de probabilité. Leur utilisation en traitement d'images et plus particulièrement en recalage d'images, est en revanche marginale. En effet, comme nous l'avons vu au chapitre 1, les mesures de similarité classiques en recalage reposent en général sur les statistiques d'ordre 2 (covariance), et n'utilisent donc qu'une partie de l'information statistique disponible dans les images à recaler. Au cours de ce travail de thèse, nous avons pensé que les cumulants d'ordre 2 mais également d'ordres supérieurs pourraient permettre de construire une nouvelle mesure de similarité pour le recalage d'images multimodales. Cette dernière se voudrait plus informative que les mesures statistiques d'ordre 2 mais aussi de variance et de coût numérique d'estimation plus petits que ceux de l'IM.

3.2.1 Première ébauche

L'idée d'utiliser les cumulants semblant prometteuse, nous avons décidé de la mettre en pratique. Pour ce faire, nous avons commencé par construire une mesure de similarité basée sur une somme pondérée des cumulants marginaux d'ordre 2, 3 et 4. Cette étude, présentée dans

[Rubeaux et al., 2009], montre des résultats prometteurs pour le recalage non-rigide d'images scanner synthétiques. Nous avons donc voulu poursuivre dans cette voie, mais nous voulions trouver une manière d'agencer les cumulants entre eux qui permette d'établir un lien direct avec l'IM. Nous nous sommes alors intéressés aux développements de **Gram-Charlier et d'Edgeworth** qui permettent d'exprimer la densité de probabilité d'une VA en fonction de ses cumulants. Puis, nous nous sommes rapidement tournés vers les approximations de l'IM basées sur de tels développements, utilisées essentiellement en Analyse en Composantes Indépendantes (ACI) [Comon, 1994, Trappenberg et al., 2006].

3.2.2 Développements de Gram-Charlier et d'Edgeworth

Les développements de Gram-Charlier et d'Edgeworth cherchent à approximer une densité de probabilité p_x à partir de ses cumulants et des polynômes d'Hermite. Lorsqu'une VA est non gaussienne, il peut paraître intéressant de décrire les écarts au caractère gaussien. Cette idée intuitive trouve une justification mathématique dans le développement de la densité au voisinage de la loi gaussienne la plus proche. La partie non-gaussienne de la densité de probabilité est alors donnée par les cumulants.

Dans la littérature, il existe une confusion courante entre le développement d'Edgeworth et celui de Gram-Charlier. Il nous apparaît important ici de clarifier les choses. La seule différence entre le développement de Gram-Charlier et celui d'Edgeworth réside dans l'agencement des termes du développement. En effet, comme il est précisé dans [McCullagh, 1987], les premiers termes correctifs dans le développement de Gram-Charlier sont de l'ordre $O(n^{-1/2})$, $O(n^{-1})$, $O(n^{-3/2})$, $O(n^{-1})$ et n'observent donc pas une décroissance monotone. Le développement d'Edgeworth remédie à ce problème, en faisant référence au théorème de la limite centrale : il rassemble les termes qui correspondent à la même puissance (négative) de n [Lacoume et al., 1997], où n correspond au nombre de VAs indépendantes dont la somme décrit entièrement la VA considérée. Ceci permet à la fois de contrôler l'erreur commise quand la série est tronquée, mais aussi de considérer le développement d'Edgeworth comme un vrai développement asymptotique [McCullagh, 1987]. C'est pourquoi nous le préférons dans les calculs que nous menons, bien que le développement de Gram-Charlier soit aussi couramment utilisé.

Nous ne donnerons pas ici les détails du développement conduisant aux développements d'Edgeworth et de Gram-Charlier mais le lecteur peut se référer à [McCullagh, 1987], [Barndorff-Nielsen and Cox, 1989] pour une construction « classique » se basant sur les fonctions génératrices des moments et des cumulants, ou à [Kolassa, 1994] pour des constructions alternatives. Les premiers termes du développement univarié d'Edgeworth autour de la distribution normale de la densité de probabilité marginale $p_{x^{(i)}}$ est donné par [Barndorff-Nielsen and Cox, 1989] :

$$\begin{aligned} p_{x^{(i)}}(u^{(i)}) &= \phi_{x^{(i)}}(u^{(i)}) \left(1 + \frac{1}{3!} \kappa_3 H_3(u^{(i)}) + \frac{1}{4!} \kappa_4 H_4(u^{(i)}) \right. \\ &\quad \left. + \frac{10}{6!} \kappa_3^2 H_6(u^{(i)}) + \frac{1}{5!} \kappa_5 H_5(u^{(i)}) + \frac{35}{7!} \kappa_3 \kappa_4 H_7(u^{(i)}) + \dots \right) \end{aligned} \quad (3.32)$$

où $\phi_{x^{(i)}}$ est la distribution normale univariée (3.1) et H_i est le polynôme d'Hermite d'ordre i (3.4). Il est important de préciser ici que le développement de la densité de probabilité se fait autour de la distribution normale ayant les mêmes caractéristiques de premier et second ordre que les VAs considérées, c'est à dire les mêmes espérances et variances. C'est pour cela que nous utilisons l'indice $x^{(i)}$ à la fois pour la densité de probabilité et la distribution de la loi

normale. Sous l'hypothèse de VAs centrées-réduites, que nous faisons au cours de ce chapitre, la loi normale sera donc également centrée-réduite, comme décrite par l'équation (3.2).

A titre de comparaison, les premiers termes du développement univarié de Gram-Charlier autour de la distribution normale de la densité de probabilité marginale $p_{x^{(i)}}$ est donné par [McCullagh, 1987] :

$$\begin{aligned} p_{x^{(i)}}(u^{(i)}) &= \phi_{x^{(i)}}(u^{(i)}) \left(1 + \frac{1}{3!} \kappa_3 H_3(u^{(i)}) + \frac{1}{4!} \kappa_4 H_4(u^{(i)}) \right. \\ &\quad \left. + \frac{1}{5!} \kappa_5 H_5(u^{(i)}) + \frac{1}{6!} (\kappa_6 + 10\kappa_3^2) H_6(u^{(i)}) + \dots \right) \end{aligned} \quad (3.33)$$

Les deux séries sont identiques à l'infini, mais comme elles sont généralement tronquées après un certain nombre de termes, elles ne donnent alors pas le même résultat.

De la même manière, nous donnons l'expression de la densité de probabilité conjointe $p_{\mathbf{x}}$ approximée par un développement d'Edgeworth à l'ordre 4 ([Barndorff-Nielsen and Cox, 1989], p.175) :

$$\begin{aligned} p_{\mathbf{x}}(\mathbf{u}) &= \phi_{\mathbf{x}}(\mathbf{u}) \left(1 + \frac{1}{3!} \kappa^{i_1, i_2, i_3} H_{i_1, i_2, i_3}(\mathbf{u}) + \frac{1}{4!} \kappa^{i_1, i_2, i_3, i_4} H_{i_1, i_2, i_3, i_4}(\mathbf{u}) \right. \\ &\quad \left. + \frac{10}{6!} \kappa^{i_1, i_2, i_3} \kappa^{i_4, i_5, i_6} H_{i_1, i_2, i_3, i_4, i_5, i_6}(\mathbf{u}) \right) + O(N^{-3/2}) \end{aligned} \quad (3.34)$$

où $\phi_{\mathbf{x}}$ est la distribution normale bivariée (3.6), H_{i_1, \dots, i_n} est le tenseur d'Hermite d'ordre n qui se calcule à partir de (3.12) et $\kappa^{i_1, i_2, i_3} H_{i_1, i_2, i_3}(\mathbf{u})$ est une somme de termes obtenus par permutation des indices (i_1, i_2, i_3) , $i_j \in 1, 2$ dans le cas bivarié. Nous développons explicitement ces termes en annexe C (voir équations (C.34), (C.57) et (C.58)), et les démonstrations de ces développements sont disponibles dans [Barndorff-Nielsen and Cox, 1989], p.175. Encore une fois, le vecteur gaussien bivarié impliqué dans le développement (3.34) est supposé avoir la même espérance et la même matrice de covariance que celles de \mathbf{x} , et donc avoir la densité de probabilité définie par (3.7). Si nous tronquons à l'ordre 4 le développement de Gram-Charlier de la densité de probabilité conjointe $p_{\mathbf{x}}$, nous obtenons :

$$p_{\mathbf{x}}(\mathbf{u}) = \phi_{\mathbf{x}}(\mathbf{u}) \left(1 + \frac{1}{3!} \kappa^{i_1, i_2, i_3} H_{i_1, i_2, i_3}(\mathbf{u}) + \frac{1}{4!} \kappa^{i_1, i_2, i_3, i_4} H_{i_1, i_2, i_3, i_4}(\mathbf{u}) \right) + O(N^{-1}) \quad (3.35)$$

et nous observons immédiatement la différence entre les deux développements. Des démonstrations de la validité du développement d'Edgeworth peuvent être trouvées dans [Cramer, 1937, Feller, 1971] par exemple.

3.2.3 Analyse en Composantes Indépendantes et Information Mutuelle

Avec les développements précédents, nous avons donc accès aux approximations des densités de probabilité marginales et conjointes de nos VAs. Il apparaît alors intéressant de calculer les entropies marginales et conjointes de Shannon à partir de ces développements. Cette idée est née en traitement du signal, et plus particulièrement en Analyse en Composantes Indépendantes (ACI) [Comon, 1994], très populaire en séparation aveugle de sources. L'illustration classique de l'ACI est le problème de la soirée cocktail : lors d'une telle soirée, un nombre N de microphones enregistrent les voix des P invités. Chaque microphone

enregistre une superposition pondérée des discours des P personnes. Le problème consiste à séparer les voix les unes des autres uniquement à partir des N observations et ce sans information a priori sur les locuteurs, excepté l'indépendance de leurs messages. Au Laboratoire Traitement du Signal et de l'Image, l'ACI est utilisée en électrophysiologie humaine comme prétraitement afin de débruiter l'activité électrique de surface d'éventuels artefacts d'origine oculaire ou musculaire en exploitant l'indépendance physiologique des activités électriques à séparer [Kachenoura et al., 2008, Albera et al., 2010a, Albera et al., 2010b]. C'est en échangeant avec les collègues qui travaillent sur l'ACI, et pour qui la notion d'IM était familière, qu'est venue l'idée d'appliquer certaines mesures de dépendance utilisées en ACI au recalage. Cependant, le principe de l'ACI peut être vu comme exactement inverse à celui du recalage. En effet, le but de l'ACI est de minimiser les dépendances des VAs mises en jeu, tandis que le but du recalage est au contraire de maximiser cette dépendance. De cette différence naissent des problèmes méthodologiques sur lesquels nous revenons dans ce qui suit.

Dans [Comon, 1994], une approximation de la néguentropie marginale d'une VA est donnée, basée sur un développement d'Edgeworth tronqué de la densité de probabilité associée. Rappelons que la néguentropie d'un vecteur aléatoire \mathbf{x} est définie par :

$$J(\mathbf{x}) = H(\mathbf{x}_g) - H(\mathbf{x}) \quad (3.36)$$

où $H(\mathbf{x}_g)$ est l'entropie d'un vecteur aléatoire gaussien de mêmes espérance et matrice de covariance que celle de \mathbf{x} . Notons que pour des statistiques d'ordre 1 et 2 données, la loi normale est celle engendrant une entropie maximale [Lacoume et al., 1997]. La néguentropie traduit quant à elle le déficit d'entropie d'une loi vis à vis de la loi normale la plus proche. Dans [Comon, 1994], l'IM est réécrite en fonction de la néguentropie :

$$IM(\mathbf{x}) = J(\mathbf{x}) - \sum_{i=1}^N J(x^{(i)}) + \frac{1}{2} \log \left(\frac{\prod_{i=1}^N \kappa_{\mathbf{x}}^{i,i}}{|\kappa_{\mathbf{x}}|} \right) \quad (3.37)$$

En blanchissant les observations, c'est-à-dire en appliquant une transformation linéaire au vecteur \mathbf{x} de manière à obtenir un vecteur aléatoire de matrice de covariance égale à la matrice identité, une transformation orthogonale résolvant le problème d'ACI posé peut être calculée en minimisant le second terme du membre de droite de (3.37) pendant que les deux autres termes restent constants. Il n'est donc pas nécessaire d'approximer l'IM pour résoudre le problème d'ACI, une approximation des néguentropies marginales suffit [Comon, 1994]. Cependant, cette méthode n'est pas directement applicable à notre problème. Nous ne détaillerons pas les développements mathématiques nécessaires, car ceux-ci ne permettent pas de construire une IM adéquate pour notre propos. En effet, dans le cadre du recalage d'images, aucune transformation préalable des images ne permettra de rendre constante l'entropie conjointe des deux images vis à vis de la transformation recherchée, d'où la nécessité d'approximer également l'entropie conjointe. Néanmoins, les travaux de [Comon, 1994] méritent l'attention du lecteur de par leur originalité, qui peut également se référer à [Lacoume et al., 1997, Hyvarinen et al., 2001] pour de plus amples détails.

Nous avons continué notre étude bibliographique des approximations de l'IM basées sur les cumulants et le développement d'Edgeworth ou de Gram-Charlier qui pourraient s'adapter à notre problème. En particulier, nous avons cherché des développements répondant à la formulation classique de l'IM, que nous rappelons ici :

$$IM(\mathbf{x}) = H(x^{(1)}) + H(x^{(2)}) - H(\mathbf{x}) \quad (3.38)$$

avec $H(\mathbf{x})$ l'entropie conjointe et $H(x^{(1)})$, $H(x^{(2)})$ les entropies marginales des VAs (les images à recalcr). Nous voulions donc nous doter à la fois d'une approximation de l'entropie marginale,

mais aussi de l'entropie conjointe basée sur un développement tronqué de type Gram-Charlier ou Edgeworth. Nous nous sommes alors intéressés à l'approximation de l'entropie marginale donnée par [Amari et al., 1996], toujours dans le contexte de l'ACI, par un développement de Gram-Charlier à l'ordre 4. Celle-ci est donnée par :

$$H(x) \approx 0.5 \log(2\pi e) - \frac{1}{2} \frac{\kappa_3^2}{3!} - \frac{1}{2} \frac{\kappa_4^2}{4!} + \frac{5}{8} \kappa_3^2 \kappa_4 + \frac{1}{16} \kappa_4^3 \quad (3.39)$$

Après avoir redémontré ce résultat, en utilisant la méthode indiquée dans [Amari et al., 1996], nous nous sommes rendus compte d'une erreur dans cette formule (3.39). En effet, l'entropie marginale s'exprime plutôt par :

$$H(x) \approx 0.5 \log(2\pi e) - \frac{1}{2} \frac{\kappa_3^2}{3!} - \frac{1}{2} \frac{\kappa_4^2}{4!} + \frac{3}{8} \kappa_3^2 \kappa_4 + \frac{1}{16} \kappa_4^3 \quad (3.40)$$

où $\frac{3}{8}$ remplace $\frac{5}{8}$.

Nous avons ensuite cherché une estimation de l'entropie conjointe. Dans le contexte de l'ACI, l'article [Akaho, 1999] définit une approximation de l'entropie conjointe par un développement de Gram-Charlier bivarié. Nous allons exposer ici le principe général de cette approximation.

Tout d'abord, les données sont centrées-réduites pour que les VAs aient une espérance nulle et une variance unité. Ceci n'est pas un problème dans notre contexte applicatif car c'est un prétraitement classique en traitement d'images. Ensuite, pour simplifier encore le développement, une étape de décorrélation des VAs mises en jeu est nécessaire. Ce prétraitement correspond à l'étape de blanchiment mentionnée précédemment et utilisée dans [Comon, 1994] pour se prémunir de l'utilisation de la néguentropie (et donc de l'entropie) conjointe. Nous notons $x^{(1)}$ et $x^{(2)}$ les VAs centrées-réduites et $x_b^{(1)}$ et $x_b^{(2)}$ les VAs centrées-réduites blanchies, l'étape de décorrélation consiste à déterminer une matrice \mathbf{M} telle que la matrice de covariance de $\mathbf{x}_b = (x_b^{(1)}, x_b^{(2)})^\top = \mathbf{M}(x^{(1)}, x^{(2)})^\top$ soit l'identité. On cherche alors \mathbf{M} comme étant l'inverse d'une racine carrée de la matrice de covariance de $\mathbf{x} = (x^{(1)}, x^{(2)})^\top$. Mathématiquement, cela peut s'écrire :

$$\mathbf{x}_b = \mathbf{M}\mathbf{x} = \begin{pmatrix} c^+ & c^- \\ c^- & c^+ \end{pmatrix} \mathbf{x} \quad (3.41)$$

avec :

$$c^+ = \frac{(1 + \rho)^{-\frac{1}{2}} + (1 - \rho)^{-\frac{1}{2}}}{2} \quad (3.42)$$

et :

$$c^- = \frac{(1 + \rho)^{-\frac{1}{2}} - (1 - \rho)^{-\frac{1}{2}}}{2} \quad (3.43)$$

où $\rho = \mathbb{E}[x^{(1)}x^{(2)}]$ est la covariance entre $x^{(1)}$ et $x^{(2)}$ présentée dans la section 3.1.1.2. Une approximation de l'entropie conjointe est ensuite calculée à partir de ces VAs blanchies $x_b^{(1)}$ et $x_b^{(2)}$, et de l'approximation (3.35) de la densité de probabilité conjointe par un développement de Gram-Charlier à l'ordre 4. Elle est donnée par :

$$\begin{aligned}
H(\mathbf{x}_b) &\approx 1 + \log(2\pi) - \frac{1}{12} \left(\kappa^{1,1,1^2} + \kappa^{2,2,2^2} + 3\kappa^{1,1,2^2} + 3\kappa^{1,2,2^2} \right) \\
&\quad - \frac{1}{48} \left(\kappa^{1,1,1,1^2} + 4\kappa^{1,1,1,2^2} + 6\kappa^{1,1,2,2^2} + 4\kappa^{1,2,2,2^2} + \kappa^{2,2,2,2^2} \right)
\end{aligned} \tag{3.44}$$

où les cumulants sont calculés à partir des variables blanchies, en utilisant l'équation (3.30) et en posant $\rho = 0$. Ensuite, l'approximation de l'entropie conjointe des VAs originales $x^{(1)}$ et $x^{(2)}$ est déduite en exploitant la relation suivante :

$$H(\mathbf{x}) = H(\mathbf{x}_b) + \frac{1}{2} \log(1 - \rho^2) \tag{3.45}$$

Notre recherche bibliographique nous a finalement mené vers [Van Hulle, 2005a], qui propose une approximation de l'entropie conjointe à partir d'un développement d'Edgeworth tronqué à l'ordre 3 (c'est à dire que dans l'équation (3.34), seuls les deux premiers termes entre parenthèses sont conservés). Dans cet article, il n'est pas fait explicitement allusion au blanchiment des données avant le calcul. Cependant, si l'on prend la dernière ligne de l'équation (2) de l'article [Van Hulle, 2005a] et qu'on la réécrit dans le cas bivarié (c'est à dire pour $d = 2$ d'après sa notation), l'entropie conjointe est approximée par :

$$\begin{aligned}
H(\mathbf{x}) &\approx H(\phi_{\mathbf{x}}(\mathbf{u})) - \frac{1}{12} \left(\sum_{i=1}^2 \kappa^{i,i,i^2} + 3 \sum_{i,j=1, i \neq j}^2 \kappa^{i,i,j^2} + \frac{1}{6} \sum_{i,j,k=1, i \neq j \neq k}^2 \kappa^{i,j,k^2} \right) \\
&= H(\phi_{\mathbf{x}}(\mathbf{u})) - \frac{1}{12} \left(\kappa^{1,1,1^2} + \kappa^{2,2,2^2} + 3\kappa^{1,1,2^2} + 3\kappa^{1,2,2^2} \right) \\
&= 1 + \log(2\pi) + 0.5 \log(|\boldsymbol{\kappa}_{\mathbf{x}}|) - \frac{1}{12} \left(\kappa^{1,1,1^2} + \kappa^{2,2,2^2} + 3\kappa^{1,1,2^2} + 3\kappa^{1,2,2^2} \right)
\end{aligned} \tag{3.46}$$

Cette approximation est égale à l'approximation (3.44) pour des données blanchies que l'on aurait tronqué à l'ordre 3 (les développements de Gram-Charlier et d'Edgeworth sont en effet équivalents à l'ordre 3), à l'exception de l'entropie de la loi normale bivariée. En effet, le terme $1 + \log(2\pi)$ de (3.44) correspond à l'entropie d'une loi normale bivariée blanchie, tandis que le terme $1 + \log(2\pi) + 0.5 \log(|\boldsymbol{\kappa}_{\mathbf{x}}|)$ de (3.46) laisse entendre que la loi normale bivariée pourrait ne pas être blanchie. Deux explications sont possibles. Soit l'auteur suppose implicitement que $\boldsymbol{\kappa}_{\mathbf{x}} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$, et se ramène ainsi à l'approximation de l'entropie conjointe pour des données préalablement blanchies. Dans le cas contraire, son approximation semble incorrecte car nous montrerons dans les développements à venir l'approximation de l'entropie conjointe pour des données non blanchies, qui diffère de (3.46). Quoi qu'il en soit, cet article [Van Hulle, 2005a] et les autres par ce même auteur, dont notamment [Van Hulle, 2005b, Van Hulle, 2008], montrent l'intérêt de l'IM approximée par un développement d'Edgeworth et apportent des pistes à approfondir. Tout d'abord, l'approximation faite dans l'équation (2) de [Van Hulle, 2005a] nous a été utile lorsque nous avons voulu contruire nos propres mesures de similarité, comme nous le verrons par la suite. De plus l'approximation d'une densité de probabilité par mélange de gaussiennes de [Van Hulle, 2005b] permet d'améliorer la robustesse de l'estimation de l'entropie. Nous pensons que c'est un chemin à suivre dans des développements futurs. Enfin, la possibilité d'étendre facilement cette mesure aux dimensions supérieures [Van Hulle, 2005a] est

également une piste à creuser, car comme nous l'avons vu au chapitre 2, c'est un des challenges actuels de l'estimation de l'IM, notamment pour exploiter des a priori d'ordre spatial.

A partir de ces différents articles, nous avons construit une première approximation de l'IM, comme présentée dans [Trappenberg et al., 2006], en injectant (3.45) et (3.40) dans la définition de l'IM (3.38).

3.2.4 Remarques sur cette approximation de l'IM

La nécessité d'un blanchiment qu'impose l'approximation de l'IM détaillée précédemment ne convient pas à notre méthode. En effet, en ACI, comme nous l'avons déjà dit, le but de la mesure de similarité est de minimiser la dépendance des variables considérées, tandis qu'en recalage d'images, cette même mesure de similarité permet au contraire de chercher la dépendance maximale entre les images. En pratique, cette différence nous a posé des problèmes. En effet, l'étape de blanchiment des VAs, en l'occurrence les images à recalcr, est possible au début du recalage. Mais plus on s'approche de l'optimum, c'est à dire plus nos images sont recalées et donc « semblables » au sens statistique du terme, plus elles deviennent difficiles à décorréliser. En poussant le raisonnement à l'extrême, deux VAs (nos images) centrées-réduites identiques auront une covariance $\rho = E[x^{(1)}x^{(2)}]$ avec $x^{(1)} = x^{(2)}$, soit $\rho = E[x^{(1)}x^{(1)}] = Var(x^{(1)}) = 1$. Par conséquent, les expressions (3.42) et (3.43) seront égales à $\frac{1}{2\sqrt{2}}$, ce qui donne, d'après (3.41) :

$$(x_b^{(1)}, x_b^{(2)})^\top = \frac{1}{2\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} (x^{(1)}, x^{(2)})^\top \quad (3.47)$$

l'étape de blanchiment dans ce cas précis. On voit bien ici que lorsque les VAs supposées blanches sont à nouveau identiques et donc corrélées, le blanchiment par cette méthode n'est donc pas possible. Mais quelle que soit la technique employée, le blanchiment de deux VAs centrées-réduites totalement corrélées (i.e. $\rho = 1$) est impossible. Il n'existe aucune transformation linéaire capable d'une telle action. Nous avons tout de même réalisé quelques tests préliminaires avec cette approximation de l'IM basée sur un développement de Gram-Charlier pour des VAs centrées-blanches, et les résultats se sont avérés en adéquation avec ce que laissait supposer la théorie. En effet, sur le test synthétique donné sur la figure (3.1), nous avons comparé l'IM classique estimée par un estimateur à histogramme et l'approximation de l'IM basée sur un développement de Gram-Charlier à l'ordre 4 comme décrit dans [Trappenberg et al., 2006], entre deux images scanner pour une translation entre ces deux images allant de 0 à 70 pixels. On voit tout de suite que la version classique atteint bien son maximum à l'optimum (c'est à dire pour une translation nulle entre les 2 images), alors que l'IM approximée par le développement de Gram-Charlier est au maximum pour une translation d'environ 5 pixels, avant de redescendre vers 0 à l'optimum réel.

Cette approximation n'est donc clairement pas adaptée à notre contexte, puisqu'une telle mesure manquerait systématiquement l'optimum. Néanmoins, on voit sur la figure (3.1) qu'elle se comporte globalement bien, du point de vue de la régularité et de l'absence de minima locaux, par rapport à l'IM classique estimée par histogramme.

3.3 Calcul de nouvelles mesures basées sur un développement d'Edgeworth

Partant de ce constat, nous avons décidé de calculer notre approximation de l'IM ne nécessitant aucun blanchiment des VAs mises en jeu. Ceci nous a conduit à un nouveau développement

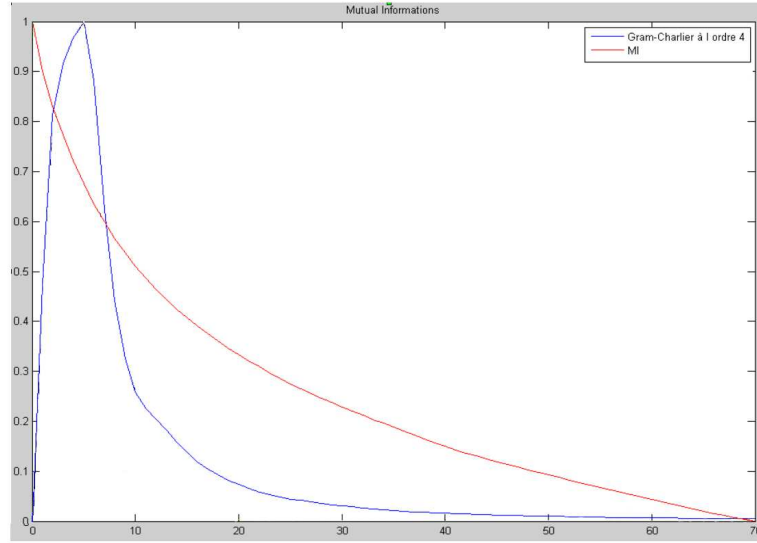


FIGURE 3.1 – Comparaison entre l'IM estimée par histogramme et celle approximée en utilisant un développement de Gram-Charlier à l'ordre 4 sur un test synthétique. On calcule ces 2 mesures entre 2 images scanner identiques pour une translation croissante de 0 à 70 pixels. Avec l'IM estimée par histogramme, le maximum se trouve bien pour une translation nulle, ce qui n'est pas le cas pour l'approximation de l'IM considérée.

de l'entropie conjointe. Par souci d'homogénéité par rapport aux approximations utilisées pour ce calcul de l'entropie conjointe, nous avons également calculé une nouvelle approximation de l'entropie marginale.

3.3.1 Nouvelles hypothèses de travail

Le développement utilisé dans les approximations précédentes (3.45) et (3.40) est un développement de Gram-Charlier, et nous avons vu que celui-ci n'était pas un estimateur asymptotiquement vrai de la densité de probabilité [McCullagh, 1987]. C'est pourquoi nous lui avons tout d'abord préféré un développement d'Edgeworth. Ensuite, la seule hypothèse initiale que nous avons conservé sur nos VAs est qu'elles soient centrées-réduites, puisque ce n'est pas un souci dans un contexte de recalage. Un simple pré-traitement sur les valeurs des niveaux de gris des images à recaler permet de centrer et normaliser nos variables aléatoires.

Partant de ces hypothèses, et contrairement aux approximations de la littérature [Akaho, 1999, Van Hulle, 2005a], il est nécessaire de prendre en compte la corrélation entre les VAs $x^{(1)}$ et $x^{(2)}$. Cela se traduit en pratique par l'apparition de la matrice de covariance de \mathbf{x} , que nous avons donné dans l'équation (3.8). Nous avons donc mené les calculs menant à ces nouvelles approximations, et ils sont donnés en annexe C. Les nouvelles mesures issues de ce travail sont i) une approximation de l'entropie marginale basée sur un développement d'Edgeworth à l'ordre 3 :

$$H(x) \approx 0.5 \log(2\pi e) - \frac{1}{2} \frac{\kappa_3^2}{3!} \quad (3.48)$$

ii) une approximation de l'entropie marginale basée sur un développement d'Edgeworth à l'ordre 4 :

$$H(x) \approx 0.5 \log(2\pi e) - \frac{1}{2} \frac{\kappa_3^2}{3!} - \frac{1}{2} \frac{\kappa_4^2}{4!} - \frac{5}{72} \kappa_3^4 \quad (3.49)$$

iii) une approximation de l'entropie conjointe basée sur un développement d'Edgeworth à l'ordre 3 :

$$\begin{aligned} H_{3E}(x) = & 1 + \log(2\pi) + 0.5 \log(1 - \rho^2) \\ & - \frac{1}{12(1 - \rho^2)^3} \left(\kappa^{1,1,1^2} + \kappa^{2,2,2^2} + 3\kappa^{1,1,2^2} + 3\kappa^{1,2,2^2} \right. \\ & \quad - 6\rho(\kappa^{1,1,1}\kappa^{1,1,2} + \kappa^{1,2,2}\kappa^{2,2,2} + 2\kappa^{1,1,2}\kappa^{1,2,2}) \\ & \quad + 6\rho^2(\kappa^{1,1,2^2} + \kappa^{1,2,2^2} + \kappa^{1,1,1}\kappa^{1,2,2} + \kappa^{1,1,2}\kappa^{2,2,2}) \\ & \quad \left. - 2\rho^3(\kappa^{1,1,1}\kappa^{2,2,2} + 3\kappa^{1,1,2}\kappa^{1,2,2}) \right) \end{aligned} \quad (3.50)$$

iv) une approximation de l'entropie conjointe basée sur un développement d'Edgeworth à l'ordre 4 :

$$\begin{aligned} H_{4E}(x) = & H_{3E}(x) - \frac{1}{48(1 - \rho^2)^4} \left(\kappa^{1,1,1,1^2} + 4\kappa^{1,1,1,2^2} + 6\kappa^{1,1,2,2^2} + 4\kappa^{1,2,2,2^2} + \kappa^{2,2,2,2^2} \right. \\ & - 8\rho \left(\kappa^{1,1,1,1}\kappa^{1,1,1,2} + 3\kappa^{1,1,1,2}\kappa^{1,1,2,2} + 3\kappa^{1,1,2,2}\kappa^{1,2,2,2} + \kappa^{1,2,2,2}\kappa^{2,2,2,2} \right) + 12\rho^2 \left(\kappa^{1,1,1,2^2} \right. \\ & + 2\kappa^{1,1,2,2^2} + \kappa^{1,2,2,2^2} + \kappa^{1,1,1,1}\kappa^{1,1,2,2} + 2\kappa^{1,1,1,2}\kappa^{1,2,2,2} + \kappa^{1,1,2,2}\kappa^{2,2,2,2} \left. \right) - 8\rho^3 \left(\kappa^{1,1,1,1}\kappa^{1,2,2,2} \right. \\ & + 3\kappa^{1,1,1,2}\kappa^{1,1,2,2} + \kappa^{1,1,1,2}\kappa^{2,2,2,2} + 3\kappa^{1,1,2,2}\kappa^{1,2,2,2} \left. \right) + 2\rho^4 \left(3\kappa^{1,1,2,2^2} + \kappa^{1,1,1,1}\kappa^{2,2,2,2} \right. \\ & \left. + 4\kappa^{1,1,1,2}\kappa^{1,2,2,2} \right) \left. \right) \\ & - \frac{10}{144(1 - \rho^2)^6} \left(\kappa^{1,1,1,1^4} + 6\kappa^{1,1,1^2}\kappa^{1,1,2^2} + 15\kappa^{1,1,1^2}\kappa^{1,2,2^2} + 20\kappa^{1,1,1^2}\kappa^{2,2,2^2} + 15\kappa^{1,1,1^2}\kappa^{2,2,2^2} \right. \\ & + 6\kappa^{1,2,2^2}\kappa^{2,2,2^2} + \kappa^{2,2,2^4} - 12\rho \left(\kappa^{1,1,1^3}\kappa^{1,1,2} + 10\kappa^{1,1,1^2}\kappa^{1,2,2}\kappa^{2,2,2} + 5\kappa^{1,1,1^2}\kappa^{1,1,2}\kappa^{1,2,2} \right. \\ & + 10\kappa^{1,1,1}\kappa^{1,1,2}\kappa^{2,2,2^2} + 5\kappa^{1,1,2}\kappa^{1,2,2}\kappa^{2,2,2^2} + \kappa^{1,2,2}\kappa^{2,2,2^3} \left. \right) + 30\rho^2 \left(\kappa^{1,1,1^2}\kappa^{1,1,2^2} + 4\kappa^{1,1,1^2}\kappa^{1,2,2^2} \right. \\ & + 6\kappa^{1,1,1^2}\kappa^{2,2,2^2} + 4\kappa^{1,1,2^2}\kappa^{2,2,2^2} + \kappa^{1,2,2^2}\kappa^{2,2,2^2} + \kappa^{1,1,1^3}\kappa^{1,2,2} + 4\kappa^{1,1,1^2}\kappa^{1,1,2}\kappa^{2,2,2} \\ & + 6\kappa^{1,1,1}\kappa^{1,1,2}\kappa^{1,2,2}\kappa^{2,2,2} + 4\kappa^{1,1,1}\kappa^{1,2,2}\kappa^{2,2,2^2} + \kappa^{1,1,2}\kappa^{2,2,2^3} \left. \right) - 40\rho^3 \left(\kappa^{1,1,1^3}\kappa^{2,2,2} \right. \\ & + 9\kappa^{1,1,1^2}\kappa^{1,2,2}\kappa^{2,2,2} + 3\kappa^{1,1,1^2}\kappa^{1,1,2}\kappa^{1,2,2} + 3\kappa^{1,1,1}\kappa^{1,1,2^2}\kappa^{2,2,2} + 9\kappa^{1,1,1}\kappa^{1,1,2}\kappa^{2,2,2^2} \\ & + 3\kappa^{1,1,1}\kappa^{1,2,2^2}\kappa^{2,2,2} + 3\kappa^{1,1,2}\kappa^{1,2,2}\kappa^{2,2,2^2} + \kappa^{1,1,1}\kappa^{2,2,2^3} \left. \right) + 30\rho^4 \left(3\kappa^{1,1,1^2}\kappa^{1,2,2^2} + 6\kappa^{1,1,1^2}\kappa^{2,2,2^2} \right. \\ & + 3\kappa^{1,1,2^2}\kappa^{2,2,2^2} + 5\kappa^{1,1,1^2}\kappa^{1,1,2}\kappa^{2,2,2} + 10\kappa^{1,1,1}\kappa^{1,1,2}\kappa^{1,2,2}\kappa^{2,2,2} + 5\kappa^{1,1,1}\kappa^{1,2,2}\kappa^{2,2,2^2} \left. \right) \\ & - 12\rho^5 \left(5\kappa^{1,1,1}\kappa^{1,1,2^2}\kappa^{2,2,2} + 11\kappa^{1,1,1}\kappa^{1,1,2}\kappa^{2,2,2^2} + 11\kappa^{1,1,1^2}\kappa^{1,2,2}\kappa^{2,2,2} + 5\kappa^{1,1,1}\kappa^{1,2,2^2}\kappa^{2,2,2} \right) \\ & \left. + 2\rho^6 \left(11\kappa^{1,1,1^2}\kappa^{2,2,2^2} + 21\kappa^{1,1,1}\kappa^{1,1,2}\kappa^{1,2,2}\kappa^{2,2,2} \right) \right) \end{aligned} \quad (3.51)$$

et v) deux approximations de l'IM basées sur un développement d'Edgeworth pour l'une à l'ordre 3, et pour l'autre à l'ordre 4, notées respectivement 3EIM et 4EIM, qui sont obtenues en

injectant (3.48)/(3.49) et (3.50)/(3.51) dans la définition de l'IM (2.9) donnée au chapitre 2. Enfin, nous avons également construit 2 approximations de l'IMN, appelées 3EIMN et 4EIMN en injectant les mêmes équations dans la définition de l'IMN (2.30) du chapitre 2.

Comme nous l'avons vu dans les chapitres précédents, l'entropie conjointe en tant que telle a été utilisée en recalage d'images médicales. Nous pouvons donc considérer que nous disposons de 6 nouvelles mesures de similarité : les entropies conjointes, les IMs et les IMNs construites à partir de développements d'Edgeworth à l'ordre 3 et 4. Nous avons testé ces mesures et les avons comparé aux mesures de similarité classiques utilisées en recalage d'images médicales multimodales. Les résultats de tous ces tests font l'objet du chapitre 4.

3.4 Conclusion

Au cours de ce chapitre, nous avons présenté les cumulants, des statistiques d'ordre supérieur qui permettent de caractériser finement une distribution de probabilité. Puis nous avons montré que les développements d'Edgeworth et de Gram-Charlier permettent d'approximer une densité de probabilité à partir de ces mêmes cumulants. En faisant un tour d'horizon bibliographique, nous avons ensuite montré qu'il était possible de dériver une expression de l'IM à partir de ces développements asymptotiques. Malheureusement, ces approximations de l'IM issues du traitement du signal ne peuvent servir notre propos à cause des hypothèses fortes qu'elles imposent sur les VAs considérées. Aussi, nous avons décidé de développer nos propres approximations qui s'affranchissent de ces hypothèses. Il en résulte 6 nouvelles mesures de similarité qui peuvent théoriquement être utilisées pour le recalage d'images médicales multimodales.

Dans le chapitre suivant, nous menons une évaluation aussi exhaustive que possible de ces nouvelles mesures de similarité et les comparons aux mesures de similarité classiques.

Bibliographie

- [Abramowitz and Stegun, 1964] Abramowitz, M. and Stegun, I. A. (1964). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*. Dover, New York.
- [Akaho, 1999] Akaho, S. Kiuchi, Y. U. S. (1999). MICA : multimodal independent component analysis. In *International Joint Conference on Neural Networks*, volume 2, pages 927–932.
- [Albera et al., 2010a] Albera, L., Comon, P., Parra, L. C., Karfoul, A., Kachenoura, A., and Senhadji, L. (2010a). Biomedical applications. In COMON, P. and JUTTEN, C., editors, *Handbook of blind source separation*. Academic Press.
- [Albera et al., 2010b] Albera, L., Kachenoura, A., Wendling, F., Senhadji, L., and Merlet, I. (2010b). A joint use of second, third and fourth order statistics for independent component analysis of biomedical signals. *EMBC 10, 32nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*.
- [Amari et al., 1996] Amari, S., Cichocki, A., and Yang, H. H. (1996). A new learning algorithm for blind signal separation. In *Advances in Neural Information Processing Systems*, pages 757–763. MIT Press.
- [Barndorff-Nielsen and Cox, 1989] Barndorff-Nielsen, O. and Cox, D. (1989). *Asymptotic techniques for use in statistics*. (Monographs on statistics and applied probability). Chapman and Hall.

- [Collignon et al., 1995] Collignon, A., Maes, F., Delaere, D., Vandermeulen, D., Suetens, P., and Marchal, G. (1995). Automated Multi-modality Image Registration based on Information Theory. In *Conference on Information Processing in Medical Imaging*, pages 263–274.
- [Comon, 1994] Comon, P. (1994). Independent component analysis, a new concept? *Signal Processing*, 36(3) :287–314.
- [Cramer, 1937] Cramer, H. (1937). Random variables and probability distributions. *Cambridge Tracts in Mathematics*.
- [Feller, 1971] Feller, W. (1971). *An introduction to probability theory and its applications. Vol. II*. Second edition. John Wiley and Sons Inc., New York.
- [Hyvarinen et al., 2001] Hyvarinen, A., Karhunen, J., and Oja, E. (2001). *Independent Component Analysis*. Wiley-Interscience.
- [Kachenoura et al., 2008] Kachenoura, A., Albera, L., Senhadji, L., and Comon, P. (2008). ICA : a potential tool for BCI systems. *IEEE Signal Processing Magazine*, 25 :57–68.
- [Kolassa, 1994] Kolassa, J. E. (1994). *Series approximation methods in statistics*. Springer-Verlag.
- [Lacoume et al., 1997] Lacoume, J. L., Amblard, P. O., and Comon, P. (1997). *Statistiques d'ordre supérieur pour le traitement du signal*. Collection Sciences de l'Ingénieur. Masson.
- [McCullagh, 1987] McCullagh, P. (1987). *Tensor Methods in Statistics*. Chapman and Hall.
- [Rubeaux et al., 2009] Rubeaux, M., Nunes, J., Albera, L., and Garreau, M. (2009). Dense motion estimation of the heart based on cumulants. In *Computers in Cardiology*, pages 609–612.
- [Stuart and Ord, 2006] Stuart, A. and Ord, J. K. (2006). *Kendall's Advanced Theory of Statistics, Vol. 1 : Distribution Theory (6th ed.)*, volume 101.
- [Trappenberg et al., 2006] Trappenberg, T., Ouyang, J., and Back, A. (2006). Input variable selection : Mutual information and linear mixing measures. *IEEE Transactions on Knowledge and Data Engineering*, 18(1) :37–46. Member-Andrew Back.
- [Van Hulle, 2005a] Van Hulle, M. (2005a). Multivariate edgeworth-based entropy estimation. *IEEE Workshop on Machine Learning for Signal Processing*, pages 311–316.
- [Van Hulle, 2005b] Van Hulle, M. M. (2005b). Edgeworth-Expanded gaussian mixture density modeling. *Neural Computation*, 17(8) :1706–1714.
- [Van Hulle, 2008] Van Hulle, M. M. (2008). Constrained subspace ICA based on mutual information optimization directly. *Neural Compututution*, 20 :964–973.
- [Viola, 1995] Viola, P. (1995). *Alignement by Maximization of Mutual Information*. PhD thesis, Massachusetts Institute of Technology Artificial Intelligence Laboratory, Boston, USA.

Chapitre 4

Résultats d'évaluation des mesures

Dans ce chapitre, nous présentons les résultats des tests d'évaluation de la qualité des mesures de similarité que nous avons développées et qui ont été présentées au chapitre précédent. L'évaluation de ces mesures est un problème complexe. Tout d'abord, une mesure de similarité n'est qu'une des composantes d'une méthode de recalage, et toutes ces composantes interagissent entre-elles. C'est pourquoi dans un premier temps, nous avons choisi d'évaluer nos mesures en dehors d'un contexte de recalage, sur des données synthétiques, afin de montrer leurs qualités et défauts intrinsèques. Nous avons ensuite cherché les optimiseurs les plus efficaces à utiliser en association avec la mesure sélectionnée dans l'étape précédente. Une fois cette étape franchie, nous avons donc à notre disposition une méthode de recalage globale. Les évaluations sur le recalage ont donc ensuite pu être réalisées. Au départ, nous nous sommes concentrés sur des tâches de recalage multimodal rigide. Du fait du faible nombre de DDL des transformations rigides (cf. chapitre 1), le recalage rigide est un premier pas essentiel dans l'estimation des qualités d'une méthode de recalage. Puis, nous nous sommes intéressés au recalage non-rigide, qui est la finalité de ce travail de thèse.

Lors de toutes ces évaluations, nous avons voulu comparer nos mesures de similarité avec les mesures de référence utilisées en recalage. Nous avons choisi les méthodes originelles, c'est à dire l'IMN estimée par histogramme [Collignon et al., 1995, Studholme et al., 1997] notée IMN_H , l'IM estimée par fenêtrage de Parzen avec noyau gaussien (IM_G) [Viola, 1995] et noyau B-Spline (IM_{BS}) [Thévenaz and Unser, 1997, Mattes et al., 2003], ainsi que l'IMN à noyau B-Spline (IMN_{BS}) [Thévenaz and Unser, 1997, Mattes et al., 2003]. Ce choix suit la tendance actuelle : la majorité des auteurs proposent de nouvelles méthodes d'estimation de l'IM, comme nous l'avons vu au chapitre 3, ce qui se justifie par le fait que l'IMN [Thévenaz and Unser, 1997, Mattes et al., 2003] reste encore aujourd'hui très compétitive en recalage.

Enfin, pour toutes nos mesures comme pour les IM estimées par fenêtrage de Parzen [Viola, 1995, Thévenaz and Unser, 1997], il est possible de choisir le nombre d'échantillons, c'est à dire le nombre de pixels de l'image utilisés pour l'estimation. Nous avons donc conduit dans la majorité des tests une comparaison de l'efficacité des mesures en fonction du nombre d'échantillons utilisés pour leur estimation. C'est un point important car le nombre d'échantillons influe directement sur le coût calculatoire, et une mesure qui n'a besoin que de peu d'échantillons pour une estimation correcte de l'IM est donc très prisée.

4.1 Présentation des données

Plusieurs types de données ont été utilisés au cours de ce travail de thèse. Nos mesures de similarité se voulant génériques, il nous a paru important d'utiliser des données inter- et

intra-patients, mono- et multimodales, issues de plusieurs types d'imagerie et représentant différents organes.

Dans la section 4.4, nous avons choisi d'utiliser une base de données du laboratoire utilisée dans le projet nommé « utilité de l'Imagerie Médicale dans l'Optimisation de la Pose de prothèses cardiaques implantables » (IMOP, projet 561 CITH no 04T187-188-189-190). La base de tests a été acquise à partir d'un appareil scanner Siemens Sensation comprenant 64 barrettes. Cette base, qui inclut plusieurs patients, est composée pour chacun d'entre-eux de vingt volumes représentant 20 instants du cycle cardiaque. La base de données représente donc pour chaque patient un volume tous les 5% du cycle cardiaque (entre les ondes R-R de l'électrocardiogramme). Chaque volume contient environ 300 coupes de 512×512 voxels, la distance inter-coupes est de 0.5 mm, l'épaisseur de coupe de 1mm, la résolution spatiale dans le plan de coupe de $0.35\text{mm} \times 0.35$ mm. Nous avons utilisé cette base car elle offrait une quantité de données importantes, utiles pour évaluer la valeur intrinsèque des mesures de similarité proposées sur des tests synthétiques. En effet, comme nous le verrons dans la section 4.4, pour faire des tests statistiquement intéressants, il nous était nécessaire de répéter l'expérience sur un nombre important d'images, afin de regarder le comportement moyen des différentes mesures.

Dans les sections 4.5, 4.6 et 4.7, qui traitent respectivement de l'évaluation des optimiseurs, du nombre d'échantillons à prendre en compte dans la procédure de recalage et du recalage rigide, nous avons utilisé la base RIRE [West et al., 1997] présentée en détail en annexe E. Cette base présente des volumes multimodaux intra-patients du cerveau. Plus particulièrement, elle offre les images d'un patient pour la base d'entraînement, et de dix-huit patients pour la base de test. Les modalités d'imagerie utilisées sont le scanner, l'IRM et le TEP. Nous avons choisi cette base car elle présente l'avantage de permettre une évaluation quantitative de la qualité d'un recalage, ce qui nous semblait primordial pour valider et justifier l'utilisation de nos mesures de similarité dans un cadre multimodal.

Enfin, dans les sections 4.8 qui présente les résultats de recalage non-rigide, deux bases de données ont été utilisées. La première est une base issue du laboratoire. Elle est composée de 24 volumes scanner acquis au centre Eugène Marquis de Rennes. Les 24 patients de l'étude, atteints d'un cancer de la prostate, ont passé un scanner de planification et ont ensuite été traités par radiothérapie externe. Les volumes ont une résolution de $512 \times 512 \times 140$ voxels. La taille des voxels est de 1.17×1.17 mm dans le plan de coupe et de 2 mm entre les coupes. Pour chacun de ces volumes, les organes d'intérêt (paroi rectale, prostate, vessie, vésicules séminales) ont été segmentés manuellement par un même observateur expert, selon le protocole clinique de radiothérapie. Le recalage est ici inter-sujet, en cherchant à recaler les patients les uns par rapport aux autres. La deuxième base de données est issue du challenge MICCAI 2010 nommé EMPIRE (Evaluation of Methods for Pulmonary Image Registration). Nous la présenterons plus en détail dans la section 4.8.2.

4.2 Critères d'évaluation

Afin de démontrer les qualités de nos mesures de similarité, nous nous sommes servis de plusieurs critères, qui sont liés à la base de données choisie.

Pour la base RIRE, des médecins experts ont segmenté manuellement 10 volumes d'intérêt sur chaque image. Les centres de ces volumes ont ensuite été calculés, et l'erreur de recalage est évaluée comme la distance euclidienne entre le centre du volume d'intérêt de l'image de référence et celui de l'image flottante recalée. Comme 10 volumes d'intérêt sont considérés, les critères finaux sont les erreurs moyenne, médiane et maximale sur les 10 volumes d'intérêt. Plus de détails concernant le calcul de ces erreurs sont donnés dans l'annexe E. Dans nos tests

(sections 4.5, 4.6 et 4.7), nous ne présentons que les erreurs moyennes qui nous semblent les plus représentatives. Mais les résultats complets de la section 4.7 sont disponibles à l'adresse http://www.insight-journal.org/rire/view_results.php, à mon nom.

Pour la base de données des patients atteints d'un cancer de la prostate, nous disposons de segmentations manuelles des organes d'intérêt. Cela permet de calculer un coefficient de Dice, qui reflète le recouvrement entre la segmentation S_1 de l'image de référence et celle de l'image flottante recalée S_2 . Ce coefficient est donné par :

$$\text{Dice} = \frac{2|S_1 \cap S_2|}{|S_1| + |S_2|} \quad (4.1)$$

Ce coefficient est égal à 1 quand les 2 segmentations se recouvrent parfaitement, et à 0 quand elles ne se recouvrent pas du tout. Nous expliquerons la démarche plus en détail dans la section correspondante (4.8.1), mais ce critère d'évaluation ne nous a pas complètement convaincu, comme nous le verrons par la suite, car il est spécifique à un problème donné. C'est pourquoi nous nous sommes tournés vers une autre base de données disponible en ligne.

Dans cette dernière base EMPIRE, 4 critères d'évaluation permettent de quantifier la qualité du recalage. Les méthodes employées pour cette évaluation sont référencées dans la littérature. Ces évaluations sont faites en mesurant, sur les images CT du thorax :

- l'alignement des frontières des poumons. Pour cela, les poumons ont été segmentés automatiquement en utilisant la méthode [van Rikxoort et al., 2009], puis une correction manuelle a été appliquée si nécessaire. Ensuite, les points situés à 20 mm à l'intérieur et à l'extérieur des surfaces segmentées des poumons ont été annotés respectivement Vin et Vout. Si il y a une différence entre les pixels annotés Vin et Vout recalés et ceux de la base de données, une pénalité est donnée. Le score final est donné sous la forme du pourcentage de pixels pour lesquels une pénalité a été donnée.
- l'alignement des scissures principales, qui sont des structures présentes dans les poumons et qui séparent les différents lobes pulmonaires. Elles ont été aussi segmentées automatiquement pour le besoin de l'étude, par la méthode [van Rikxoort et al., 2008]. Le score est calculé de la même manière que précédemment.
- la correspondance entre points annotés. 100 points ont été annotés sur chaque image de référence, puis la correspondance de ces points dans l'image flottante a été trouvée en employant une méthode semi-automatique [Murphy et al., 2008]. Une distance euclidienne moyenne est ensuite calculée entre les points de l'image recalée et ceux de l'image de référence.
- les singularités dans le champ de déformation dense. En calculant le déterminant du jacobien du champ de déformation, il est possible de voir si les déformations trouvées lors du recalage sont réalistes. Il est en effet possible ainsi de déterminer si il y a eu expansion ou contraction.

Une représentation de ces critères d'évaluation est donnée sur la figure 4.1. Tous les détails concernant cette base de données et les méthodes d'évaluation sont disponibles à l'adresse <http://empire10.isi.uu.nl/evaluation.php>.

4.3 Développements logiciels

Pour tester nos différents développements mathématiques, nous avons d'abord retenu le langage Matlab, qui offre l'avantage de pouvoir tester rapidement les algorithmes développés. Malheureusement, quand les volumes de données à traiter deviennent très importants, comme c'est le cas des volumes d'images médicales, l'utilisation de Matlab devient délicate dans des temps de calcul raisonnables. Tant que nous réalisons des recalages sur des images

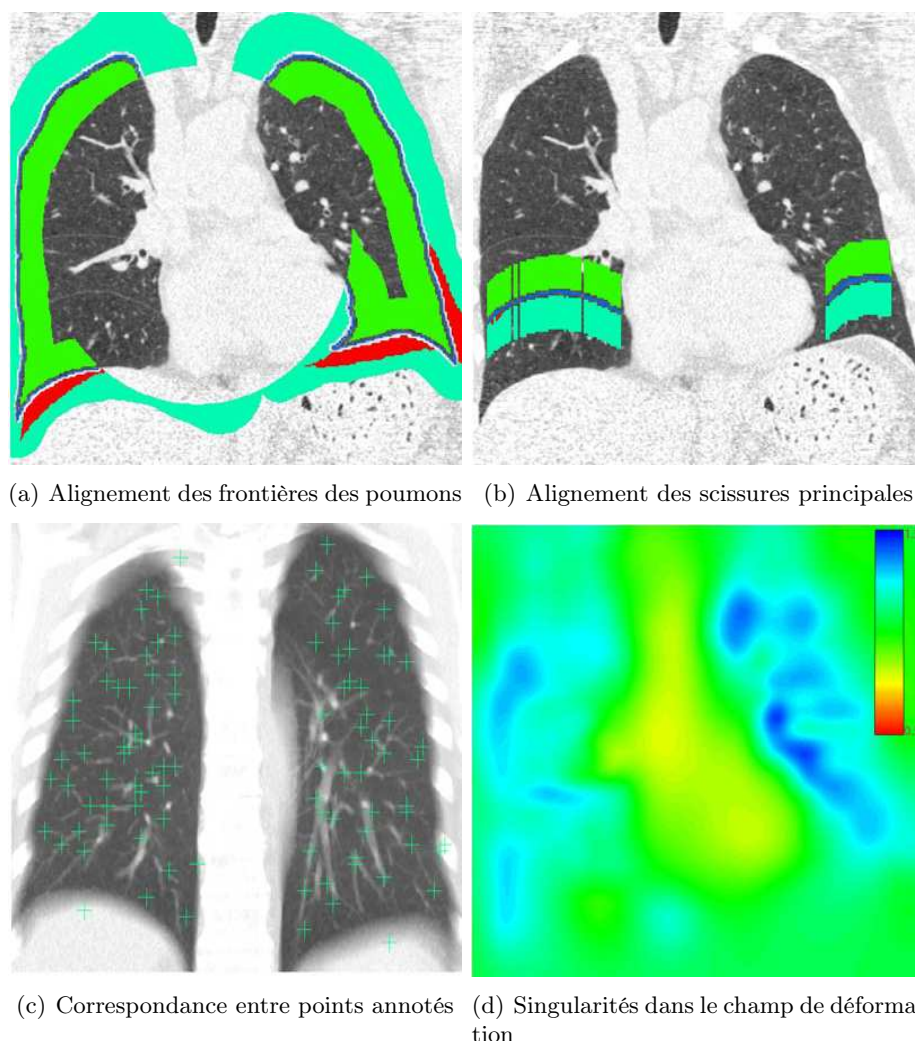


FIGURE 4.1 – Illustration des critères d'évaluation utilisés pour juger de la qualité du recalage non-rigide sur la base de données EMPIRE.

bi-dimensionnelles, cela restait envisageable, mais le passage à la dimension 3 nous a obligé à changer de stratégie. Ainsi, nous avons choisi d'utiliser la librairie C++ Insight ToolKit (ITK), qui est une librairie open-source dédiée au recalage et à la segmentation d'images. Elle offre de nombreux algorithmes et outils pour l'analyse d'images, et est disponible à l'adresse <http://www.itk.org/>. Enfin, nous avons utilisé la librairie elastiX [Klein et al., 2010] disponible à l'adresse

<http://elastix.isi.uu.nl/index.php>, qui a été développée récemment et se concentre exclusivement sur le recalage rigide et non-rigide iconique. Elle a été conçue pour le recalage d'images médicales, bien qu'elle puisse être employée pour d'autres types d'images. Cette librairie est basée sur ITK et présente des particularités intéressantes :

- les codes réalisés pour ITK sont facilement réutilisables, après quelques modifications mineures, avec la librairie elastiX.
- elle commence à être utilisée intensivement par la communauté du recalage d'images médicales, et on peut espérer qu'elle va s'enrichir, au fur et à mesure des développements des différents utilisateurs.

- enfin, et c’est sans doute le point le plus important, elle assure la reproductibilité des résultats. En effet, pour lancer un recalage avec cette librairie, il suffit de créer un fichier de paramètres qui va définir les différentes composantes de la méthode de recalage, comme donné en annexe F. Ainsi, ce fichier est « la signature » du recalage, et tout un chacun peut reproduire le même recalage, paramétré de la même manière, pour peu qu’il possède ce fichier. D’ailleurs, une liste de fichiers de paramètres associés à leurs articles est disponible sur le site de la librairie, à l’adresse http://elastix.bigr.nl/wiki/index.php/Parameter_file_database.

Nous voulons aborder ici un point important à nos yeux. Pour réaliser tous ces tests, nous nous sommes posés la question de re-coder ou non les mesures usuelles de l’IM auxquelles nous voulions comparer nos propres mesures. Lorsque l’on fait des recherches sur internet, on peut trouver un grand nombre de codes approximant l’IM, mais ils ne citent pas toujours l’article de référence correspondant à l’approximation utilisée. D’un autre côté, re-coder les méthodes de référence peut également être sujet à caution car en fonction de la mise en oeuvre, les temps de calcul peuvent grandement différer. Dans la librairie ITK, les 3 méthodes d’estimation de l’IM (IM_N , IM_G et IM_{BS}) précédemment évoquées sont disponibles et les articles sur lesquelles se basent les algorithmes sont donnés. Nous pensons que la reconnaissance de cette librairie par la communauté scientifique, tout comme sa qualité de codage permettent d’utiliser ces définitions de l’IM dans nos comparaisons. Pour s’en convaincre, nous avons examiné attentivement les codes disponibles et ils retranscrivent fidèlement les articles originaux. Nous avons donc finalement fait le choix d’utiliser les codes ITK pour les mesures de référence, ce qui assure en plus une reproductibilité de ceux-ci puisque la librairie est librement disponible. Pour ce qui est de nos propres mesures de similarité, nous nous sommes servis de la structure d’ITK qui permet de mettre en oeuvre des codes rapides. Enfin, quand après quelques modifications, ces codes ont pu être intégrés dans elastiX.

Nous avons également utilisé dans ces librairies les algorithmes codant les autres composantes d’une méthode de recalage (optimisation, modèle de déformation, interpolation, etc.). La structure même du code permet en effet de choisir chaque composante indépendamment des autres, et de les assembler pour créer une méthode globale.

4.4 Evaluation des différentes mesures de similarité

La première étape qui nous paraît essentielle pour évaluer les mesures de similarité est de regarder leur comportement général dans l’espace des paramètres de la transformation. Par cette évaluation, nous voulions voir si les mesures proposées étaient suffisamment régulières pour être utilisées dans le contexte du recalage. En effet, une mesure présentant trop d’optimums locaux sera très délicate à optimiser, et devra donc être proscrite dans ce contexte.

Pour ce faire, nous avons suivi la procédure représentée schématiquement sur la figure 4.2, en utilisant les images de la base de données IMOP. Cela consiste, tout d’abord, à sélectionner une image en 2 dimensions dans un des volumes de la base, qui sera l’image de référence. Puis, cette même image est translatée en x et y pour des valeurs comprises entre -50 et +50 mm afin d’obtenir une image flottante. Pour chaque valeur de translation, la mesure de similarité est calculée entre l’image de référence et l’image flottante, ce qui permet d’obtenir une matrice de résultats 3D. Celle-ci donne, pour chaque translation 2D, la valeur de la mesure de similarité.

Toutefois, cette expérience permet de parcourir seulement dans l’espace des paramètres de translation, c’est-à-dire dans un espace de dimension 2. Mais, nous l’avons vu, les transformations considérées pour le recalage non rigide multimodal peuvent compter plusieurs milliers de paramètres. Une étude telle que celle-ci sur un nombre important de paramètres n’est pas envisageable. Outre le fait que cela serait compliqué à représenter, les temps de calcul néces-

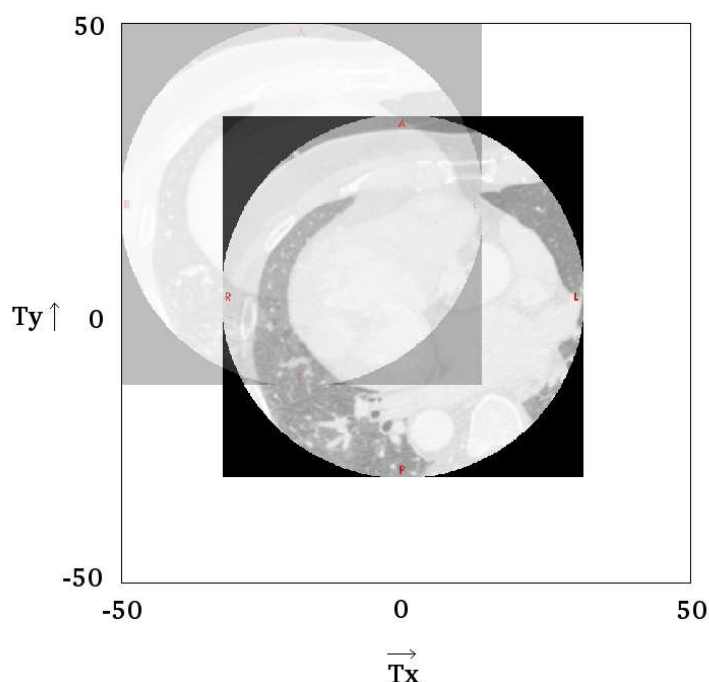


FIGURE 4.2 – Illustration des tests menés. Deux images identiques sont traduitées l’une par rapport à l’autre, et la mesure de similarité est calculée pour chaque translation.

saies seraient vraiment très importants. Toutefois, pour rendre plus réaliste la situation, nous avons décidé de réaliser ces tests en ajoutant aux 2 paramètres de translation une rotation, puis en ajoutant dans les images un bruit additif gaussien, et enfin un bruit multiplicatif de type speckle.

Au total, 10 mesures de similarité ont été soumises à ce test. Elles sont données dans le tableau 4.1. Elles correspondent aux 6 mesures que nous avons développées et présentées au chapitre précédent (3), auxquelles s’ajoutent les quatres mesures de référence précédemment citées. De plus, le nombre d’échantillons tirés de l’image utilisés pour l’estimation d’une mesure est un paramètre important, car il va grandement influencer sur la rapidité du recalage. Toutes ces mesures, à l’exception de l’IM estimée par histogramme IM_H , peuvent être estimées à partir d’un nombre réduit d’échantillons tirés aléatoirement des images. Nous avons donc réalisé ces tests en prenant en compte 50, 100, 500, 1000 et enfin 5000 échantillons issus des images.

Mesures proposées	Mesures usuelles
H_{3E}	IMN_H
H_{4E}	IM_G
$3EIM$	IM_{BS}
$4EIM$	IMN_{BS}
$3EIMN$	
$4EIMN$	

TABLE 4.1 – Mesures utilisées dans les tests de régularité.

Ces valeurs n'ont pas été choisies au hasard : dans l'article [Viola, 1995], l'auteur indique le nombre de 50 échantillons pour obtenir un recalage de qualité. Certes, le contexte applicatif est différent, mais nous allons voir à travers ces tests que ce nombre est sans doute sous évalué. Enfin, au delà de 5000 échantillons, il devient difficile d'obtenir des temps de calcul raisonnables, quelle que soit la mesure, pour une utilisation en recalage. Il faut également préciser que pour l'estimation IM_G , nous nous sommes arrêtés à 1000 échantillons, les temps de calcul devenant beaucoup trop importants pour envisager son utilisation en recalage. Nous donnerons dans la section 4.6 qui traite du recalage rigide plus de détails concernant les temps de calcul. Mais pour donner une idée, réaliser le test sur une image (ce qui correspond à $100 \times 100 = 10000$ évaluations de la mesure de similarité, puisque nous l'évaluons pour chaque translation entre -50 et +50 mm, en x et y), en prenant en compte 5000 échantillons, nécessite moins d'1 minute avec nos mesures, tout comme avec celles basées sur le fenêtrage de Parzen avec noyau B-Spline (IM_{BS} et IMN_{BS}). Pour IM_G , le même test nécessite 20 minutes.

A chaque fois, nous avons répété cette expérience sur 100 images différentes tirées au hasard dans la base de donnée IMOP et avons calculé la moyenne des mesures de similarité sur ces 100 images.

Il faut noter que nous donnons sur les différentes figures l'opposé des résultats des entropies conjointes pour une comparaison directe par rapport aux différentes IMs. En effet, l'IM atteint normalement un maximum à l'optimum, tandis que l'entropie conjointe atteint quant-à-elle un minimum. De la même façon, nous avons pris l'opposé des résultats de IM_{BS} et IMN_{BS} car par défaut dans la librairie ITK, les fonctions codant ses mesures renvoient l'opposé de l'IM.

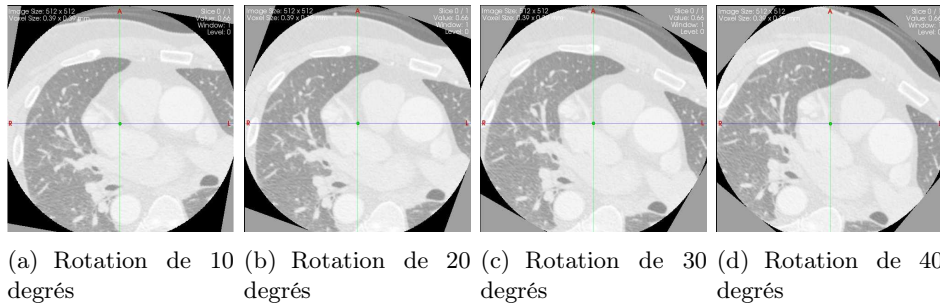


FIGURE 4.3 – Exemples d'images utilisées dans les tests, avec une rotation de 10 à 40 degrés par pas de 10 degrés.

4.4.1 Régularité pour une transformation rigide

Dans cette première série de tests, nous considérons donc une transformation rigide à 3 paramètres en 2D (2 translations et une rotation). Mis à part les paramètres dont nous avons parlé ci-avant (translations et nombre d'échantillons pris en compte dans les approximations), nous avons fait varier l'angle de rotation de 0 à 40 degrés par pas de 10 degrés. Une illustration des images générées est donnée sur la figure 4.3. Par souci de place, nous ne présentons que les résultats correspondants à une rotation de 10 degrés, mais ils sont représentatifs de tous les tests menés.

Les graphes 4.4, 4.5 et 4.6 représentent, pour une rotation de 10 degrés, les valeurs des mesures de similarité sur l'espace des translations, avec la prise en compte de respectivement 50, 100 et 500 échantillons pour l'estimation des différentes mesures de similarité.

Les simulations visant à tester les performances de l'estimation de l'IMN par histogramme IMN_H mérite un traitement à part, puisque l'on ne sélectionne pas le nombre d'échantillons

pris en compte pour son estimation, si l'on s'en réfère à l'article original [Studholme et al., 1997]. Ainsi, le résultat pour ce test initial du comportement de IMN_H est donné dans la figure 4.7, dans laquelle tous les voxels de l'image sont pris en compte dans le calcul.

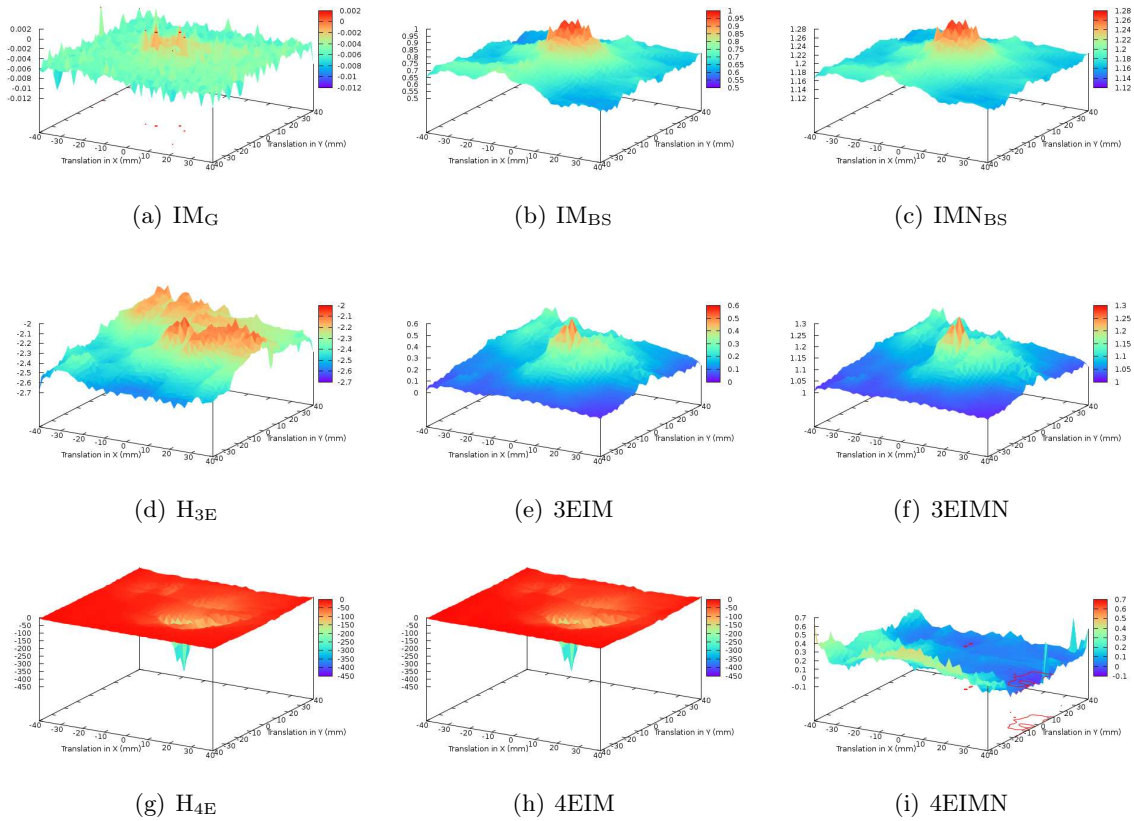


FIGURE 4.4 – Test de régularité des différentes mesures pour une transformation rigide, avec une rotation de 10 degrés, en utilisant 50 échantillons dans l'estimation des mesures. Ces figures représentent la valeur moyennée sur 100 tests des différentes mesures de similarité (en z) en fonction des translations (en x et y) considérées. Le codage va du bleu pour des valeurs faibles de la mesure, au rouge pour des valeurs importantes.

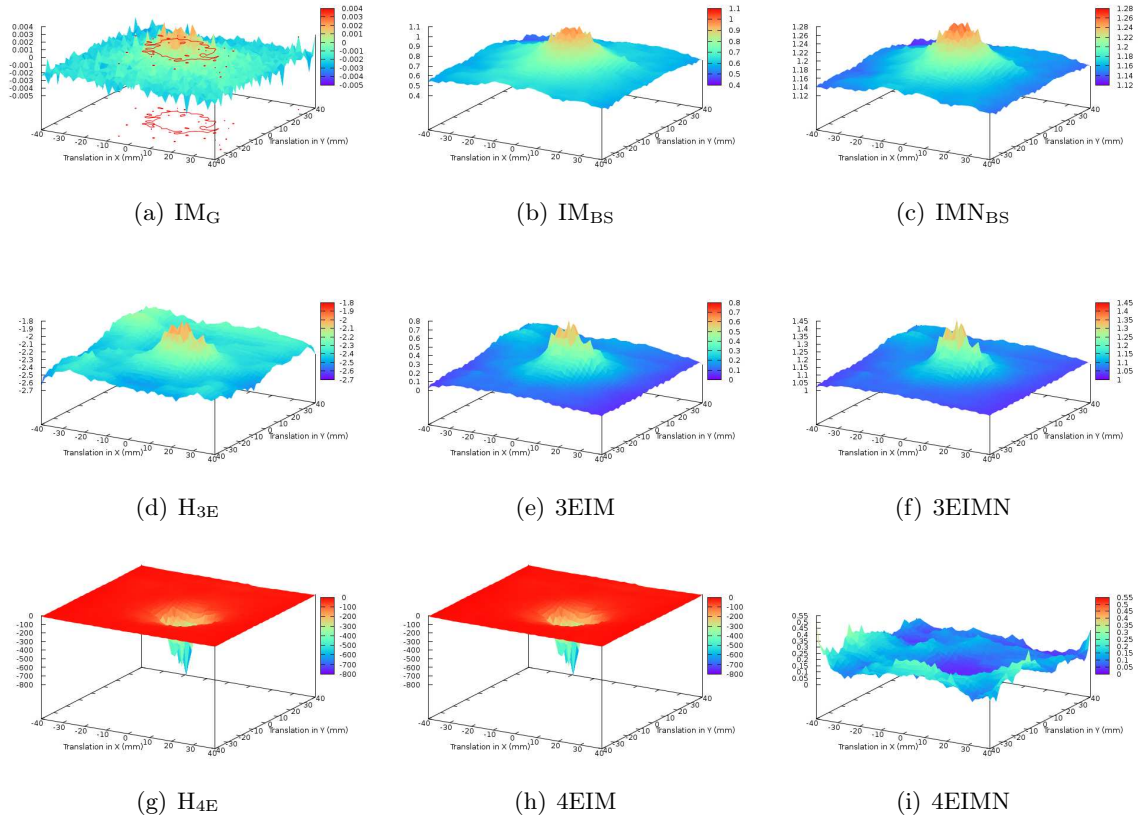


FIGURE 4.5 – Test de régularité des différentes mesures pour une transformation rigide, avec une rotation de 10 degrés, en utilisant 100 échantillons dans l'estimation des mesures. Ces figures représentent la valeur moyennée sur 100 tests des différentes mesures de similarité (en z) en fonction des translations (en x et y) considérées. Le codage va du bleu pour des valeurs faibles de la mesure, au rouge pour des valeurs importantes.

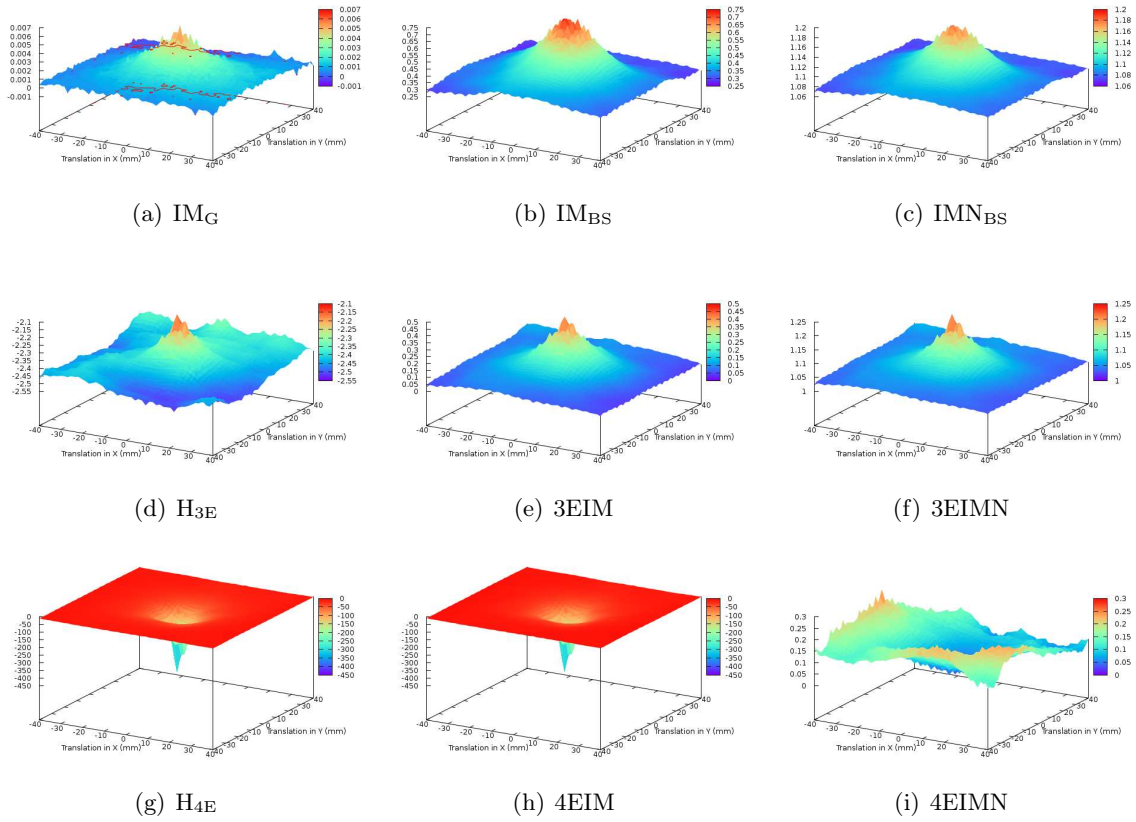


FIGURE 4.6 – Test de régularité des différentes mesures pour une transformation rigide, avec une rotation de 10 degrés, en utilisant 500 échantillons dans l'estimation des mesures. Ces figures représentent la valeur moyennée sur 100 tests des différentes mesures de similarité (en z) en fonction des translations (en x et y) considérées. Le codage va du bleu pour des valeurs faibles de la mesure, au rouge pour des valeurs importantes.

Nous pouvons tirer plusieurs enseignements des ces premiers résultats, que nous développons maintenant.

Comportement des approximations à l'ordre 4

Toutes les mesures de similarité basées par le développement d'Edgeworth à l'ordre 4 (c'est-à-dire H_{4E} , $4EIM$ et $4EIMN$) donnent le résultat opposé de ce que laissaient supposer ces approximations. En effet, les IMs approximées par ce développement ($4EIM$ et $4EIMN$) atteignent un minimum en l'optimum, alors qu'elles devraient au contraire atteindre un maximum. Ceci est notamment visible sur les figures 4.4(h), 4.5(h), 4.6(h). De la même manière, les entropies conjointes (dont nous représentons ici l'opposé pour faciliter la comparaison) devraient atteindre un minimum (et donc un maximum sur les figures) à l'optimum, ce qui est l'opposé de ce que l'on obtient sur les figures 4.4(g), 4.5(g) et 4.6(g). Ceci nous a rendu quelque peu perplexe, surtout qu'en comparaison, les résultats donnés par les mesures basées sur le développement à l'ordre 3 semblent satisfaisantes. Nous avons étudié dans le détail le comportement de ces mesures à l'ordre 4 et tâchons d'apporter ici quelques éclaircissements.

Tout d'abord, ce problème provient vraisemblablement de l'estimation de la densité de probabilité conjointe, les approximations des entropies marginales restant dans les mêmes

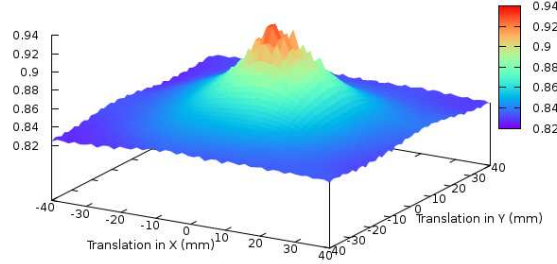


FIGURE 4.7 – Test de régularité de IMN_H pour une transformation rigide, avec une rotation de 10 degrés, en utilisant la totalité des échantillons (262000) issus de l'image. Moyenne sur 100 tests. Le codage va du bleu pour des valeurs faibles de la mesure, au rouge pour des valeurs importantes.

ordres de grandeur lorsque nous utilisons le développement à l'ordre 3 ou 4. Si nous reprenons l'équation de l'entropie conjointe basée sur un développement d'Edgeworth à l'ordre 4 (3.51) du chapitre 3, nous pouvons, pour simplifier, la réécrire comme étant :

$$H_{4E}(\mathbf{x}) = 1 + \log(2\pi) + 0.5 \log(1 - \rho^2) \quad (4.2)$$

$$- \frac{1}{12(1 - \rho^2)^3}(\dots) - \frac{1}{48(1 - \rho^2)^4}(\dots) - \frac{10}{144(1 - \rho^2)^6}(\dots)$$

où (...) contient les termes s'exprimant en fonction des cumulants, que nous ne réécrivons pas pour plus de lisibilité. Dans cette équation, les 3 termes $\frac{1}{12(1-\rho^2)^3}(\dots)$, $\frac{1}{48(1-\rho^2)^4}(\dots)$ et $\frac{10}{144(1-\rho^2)^6}(\dots)$ sont les contributions respectives des 3 termes $\frac{1}{3!}\kappa^{i_1, i_2, i_3} H_{i_1, i_2, i_3}(\mathbf{u})$, $\frac{1}{4!}\kappa^{i_1, i_2, i_3, i_4} H_{i_1, i_2, i_3, i_4}(\mathbf{u})$ et $\frac{10}{6!}\kappa^{i_1, i_2, i_3} \kappa^{i_4, i_5, i_6} H_{i_1, i_2, i_3, i_4, i_5, i_6}(\mathbf{u})$ qui apparaissent dans l'équation (3.34) exprimant la densité de probabilité conjointe $p_{\mathbf{x}}$ approximée par un développement d'Edgeworth à l'ordre 4. Quand le développement est tronqué à l'ordre 3, seul le premier de ces termes subsiste. Il nous a semblé intéressant de voir comment se comportait ces 3 termes au cours des simulations, pour essayer de mieux comprendre le comportement singulier de notre mesure tronquée à l'ordre 4. Nous avons donc isolé, dans ces tests, les valeurs prises par ces termes. Plus précisément, pour une image, et avec une rotation de 10 degrés, nous avons reporté les valeurs prises par les approximations d'ordre 4 des entropies marginales $H_{x(1)}$, $H_{x(2)}$, l'approximation 4EIM et les 3 termes que nous venons de citer dans le tableau 4.2. Nous constatons que le troisième terme croît de manière très importante au fur et à mesure que l'on s'approche de l'optimum, si bien que la valeur de l'IM à l'optimum est presque exclusivement guidée par ce terme. Ceci entraîne d'ailleurs un problème lors du calcul de l'IMN, car suivant l'équation la définissant (2.30), son estimation revient à faire la somme des entropies marginales divisée par l'entropie conjointe. Lorsque l'entropie conjointe devient très importante, il en résulte une approximation de l'IMN s'approchant de 0, comme en témoignent les figures 4.4(i), 4.5(i), 4.6(i).

Nous émettons actuellement 2 hypothèses pour expliquer ce problème :

- La première serait une erreur dans nos développements (*errare humanum est*). Nous les avons vérifiés à de maintes reprises mais cela reste possible a priori. Cependant, plusieurs arguments plaident en notre faveur. Le premier est que la mesure calculée à partir d'un

T_X	T_Y	$H_{x(1)}$	$H_{x(2)}$	$\frac{1}{12(1-\rho^2)^3}(\dots)$	$\frac{1}{48(1-\rho^2)^4}(\dots)$	$\frac{10}{144(1-\rho^2)^6}(\dots)$	4EIM
40	40	1.28	0.41	2.52	0.26	4.44	-5.53
30	30	1.28	0.52	2.48	0.29	5.61	-6.85
20	20	1.28	0.62	2.48	0.14	6.29	-7.01
10	10	1.28	0.92	2.42	0.08	17.17	-17.47
5	5	1.28	0.99	2.29	0.22	72.85	-73.08
0	0	1.28	1.04	1.94	0.21	544.12	-543.954

TABLE 4.2 – Valeurs prises par nos différentes approximations à l'ordre 4, pour une rotation de 10 degrés, en fonction des translations en X et Y.

développement d'Edgeworth tronqué à l'ordre 3 semble approximer correctement l'IM. Or, nous avons suivi exactement la même démarche pour obtenir le développement à l'ordre 4. De plus, les résultats à l'ordre 4 ne sont pas complètement aberrants, dans la mesure où les résultats, en valeur absolue, donnent un maximum à l'optimum. Il nous paraît donc peu probable qu'un calcul erroné donne des résultats si cohérents.

- La seconde serait de penser que le souci provient du développement d'Edgeworth, qui est asymptotiquement une approximation correcte de la densité de probabilité conjointe. Nous pensons que la troncature après un certain nombre de termes peut emmener une telle singularité. Pour le montrer, nous souhaitons développer l'approximation de l'entropie conjointe par un développement d'Edgeworth tronqué à l'ordre 5, afin de constater si une compensation intervient avec les termes supplémentaires issus de ce nouveau développement. Cela nécessite des développements mathématiques importants, et l'expression qui en découle ne pourra sans doute pas être appliquée au problème du recalage du fait du coût calculatoire engendré par ces nouveaux termes, mais c'est une piste à envisager.

A partir de ce constat, nous avons décidé de nous concentrer sur les mesures approximées par le développement d'Edgeworth tronqué à l'ordre 3, et c'est pourquoi les résultats à venir ne s'intéressent qu'à celles-ci. A posteriori, nous pensons que cette mesure à l'ordre 4, bien que ne pouvant pas être considérée comme une réelle approximation de l'IM, mérite tout de même notre attention. En effet, si nous nous intéressons aux résultats donnés dans les figures 4.4, 4.5 et 4.6, nous constatons que ces mesures proposent une étonnante régularité, même quand un nombre d'échantillons faible est pris en compte pour l'estimation. Pour s'en convaincre, nous donnons dans la figure 4.8 l'opposé des figures 4.4(h), 4.5(h) et 4.6(h). Nous constatons par exemple qu'en ne considérant que 50 échantillons pour l'estimation (fig.4.8(a)), cette mesure semble donner moins de maxima locaux que celles proposées sur la figure 4.4.

Comportement de l'estimation de l'IM par fenêtrage de Parzen à noyau gaussien

Nous avons constaté que dans notre contexte de recalage d'images médicales, l'IM estimée par fenêtrage de Parzen à noyau gaussien (IM_G) semble nécessiter plus de 50 échantillons pour être estimée correctement. Ce nombre d'échantillons était donné comme suffisant dans l'article présentant la méthode [Viola, 1995], mais la figure 4.4(a) montre qu'avec ce nombre d'échantillons, de nombreux optimums locaux sont présents, ce qui rend la mesure inopérante. Quand le nombre d'échantillons augmente, cette mesure devient plus régulière, mais au prix d'un coût calculatoire devenant rapidement très important par rapport aux autres méthodes.

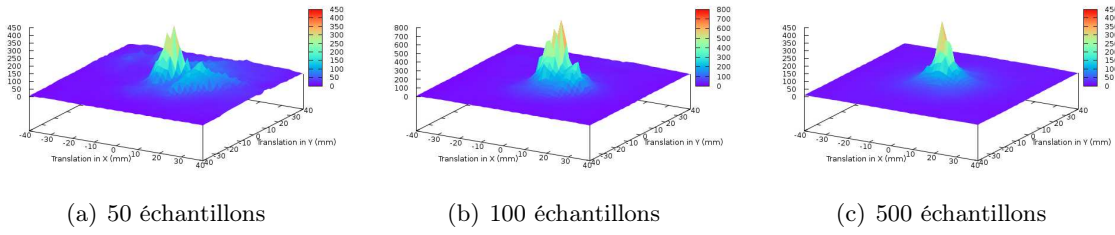


FIGURE 4.8 – Test de régularité pour une transformation rigide, avec une rotation de 10 degrés. Représentation de l’opposé de 4EIM avec la prise en compte d’un nombre croissant d’échantillons dans le calcul de l’approximation.

Comportement de l’estimation de l’IM par histogramme

Nous constatons que cette estimation se comporte bien, mais là encore au prix d’un coût calculatoire important, puisque tous les voxels (ici environ 260000) de l’image sont pris en compte dans son estimation, si l’on s’en tient à la définition originale. Certes, nous aurions pu coder la méthode d’estimation par histogramme prenant comme paramètre le nombre d’échantillons. Plusieurs points nous en ont dissuadé. Tout d’abord, si l’on s’en tient à la définition de l’article original [Collignon et al., 1995], il n’est pas fait référence à ce nombre d’échantillons. De plus, re-coder une méthode reste toujours sujet à caution. Enfin, l’IM estimée par histogramme est considérée en général comme moins efficace que ses consœurs utilisant le fenêtrage de Parzen. Cela rend également difficile la comparaison avec les autres méthodes, un des buts principaux à travers les différentes expériences menées étant justement d’étudier le comportement des mesures en fonction du nombre d’échantillons pris en compte dans leur estimation. C’est pourquoi dans la suite de ces tests préliminaires, nous ne présenterons plus les résultats pour cette mesure IMN_H .

Remarques globales Comme nous pouvions nous y attendre, augmenter le nombre d’échantillons utilisés dans les approximations de ces mesures de similarité améliore très sensiblement ces mesures, en lissant les optimums locaux. Les mesures semblent les plus performantes sur ces tests synthétiques sont l’IM et l’IMN estimée par fenêtrage de Parzen à noyau B-Splines (IM_{BS} et IMN_{BS}) ainsi que nos approximations tronquées à l’ordre 3. Au sujet de ces dernières, nous constatons que les mesures d’IM (3EIM, 3EIMN) semblent présenter moins de maxima locaux que la seule entropie conjointe H_{3E} . C’est ce que laissait supposer la théorie car comme nous l’avons vu au chapitre 1, l’IM a été utilisée en recalage d’images après l’entropie conjointe, la première cherchant à améliorer la régularité de la seconde par la prise en compte des entropies marginales des images à recalier. Nous pouvons également constater que les 2 mesures 3EIM et 3EIMN issues de nos développements donnent des résultats très similaires. Cela nous a conduit à nous concentrer, par la suite, sur la seule 3EIM. Le processus d’évaluation étant relativement long, nous avons dû faire ce choix relativement tôt. Dans le futur néanmoins, l’étude précise de 3EIMN nous semble nécessaire, au moins à titre de comparaison. Finalement, nous constatons notamment que ces mesures sont relativement régulières même lorsqu’un faible nombre d’échantillons est pris en compte. Si nous augmentons l’angle de rotation, les résultats se confirment. La figure 4.9 donne les résultats de l’expérience pour un angle de 40 degrés, avec la prise en compte de 500 échantillons. Nous constatons que dans cette situation, nos approximations à l’ordre 3 semblent même plus régulières que les IMs approximées par fenêtrage de Parzen à noyau B-Spline.

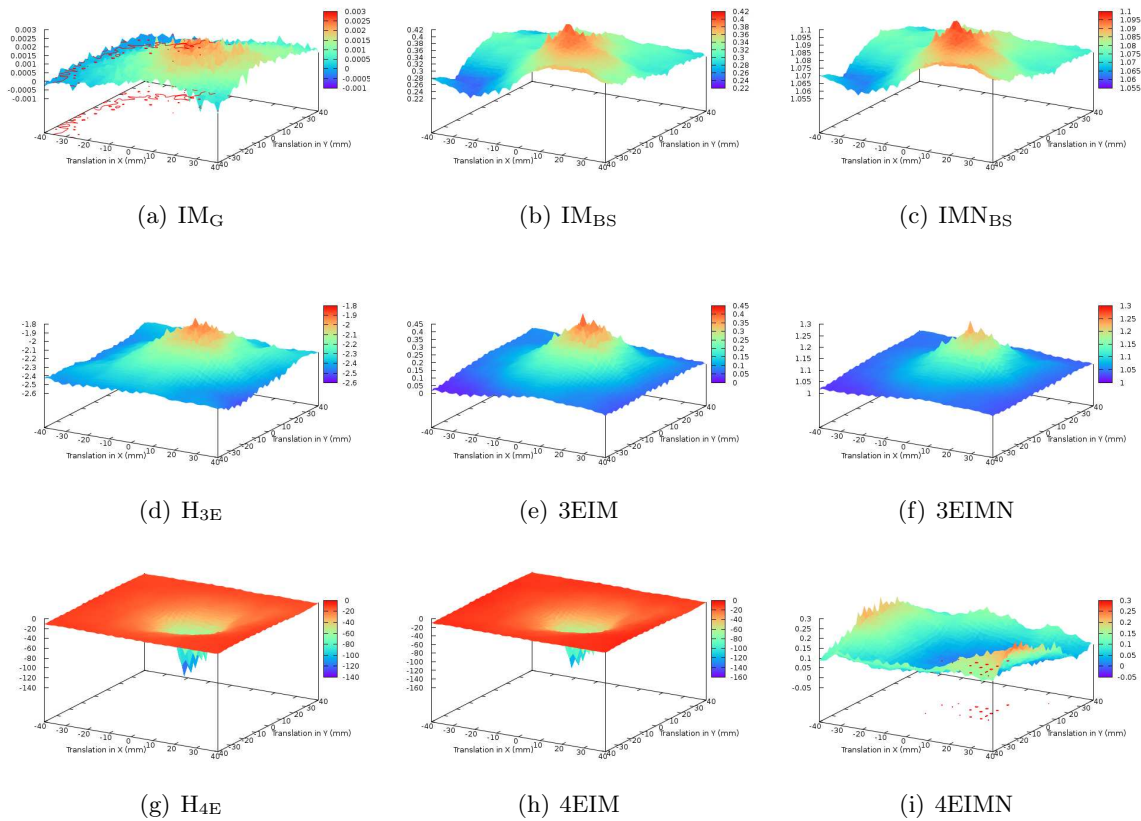


FIGURE 4.9 – Test de régularité des différentes mesures pour une transformation rigide, avec une rotation de 40 degrés, en utilisant 500 échantillons dans l'estimation des mesures. Ces figures représentent la valeur moyennée sur 100 tests des différentes mesures de similarité (en z) en fonction des translations (en x et y) considérées.

4.4.2 Régularité face au bruit

Dans ces simulations, plutôt que d'ajouter une rotation aux 2 translations, nous avons choisi d'ajouter un bruit, afin de tester la robustesse des mesures dans un environnement bruité. En effet, nous savons que les images médicales sont souvent perturbées par du bruit qui peut compliquer le recalage. Ce bruit peut avoir plusieurs origines. Il peut être dû directement à la technique d'imagerie employée, ou encore aux algorithmes de reconstruction utilisés pour créer les images 3D. Il est difficile de modéliser le bruit présent dans ces images, et cette étude ne prétend pas le reproduire fidèlement. Cependant, il nous a paru intéressant d'observer le comportement des différentes mesures dans un environnement bruité. Nous avons commencé par appliquer aux images un bruit additif gaussien d'écart type croissant. Puis, nous nous sommes intéressés au bruit multiplicatif de type speckle (ou granularité) qui est connu pour venir perturber les images ultrasonores, dont l'échographie fait partie. Ces bruits ont été générés à l'aide d'un code source disponible en ligne à l'adresse : <http://www.insight-journal.org/browse/publication/721>.

Dans la première série de tests, les images ont été dégradées par un bruit gaussien d'écart-type croissant de 1, 5, 10, 50, 100 et 500, et considérant toujours une translation de ± 50 mm en x et y . Quelques exemples d'images générées sont donnés figure 4.10. À l'œil nu, il est

très difficile de discerner la présence de ce bruit, pour un écart-type inférieur à 100. Comme pour les tests précédents, nous avons également fait varier le nombre d'échantillons. Nous ne donnons encore une fois que quelques résultats significatifs de tous les tests menés.

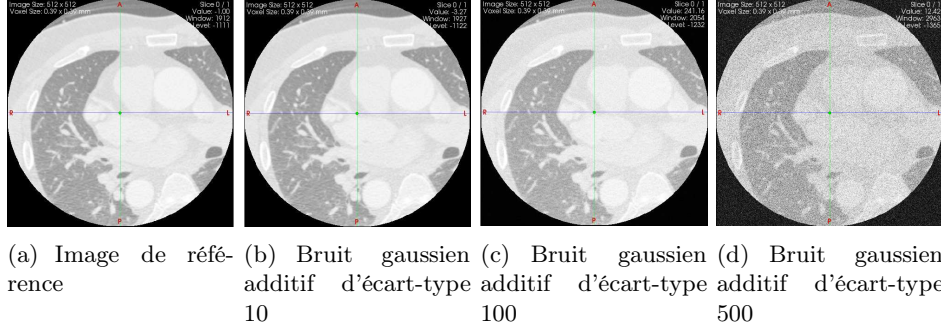


FIGURE 4.10 – Exemples d'images bruitées par un bruit gaussien additif d'écart-type croissant.

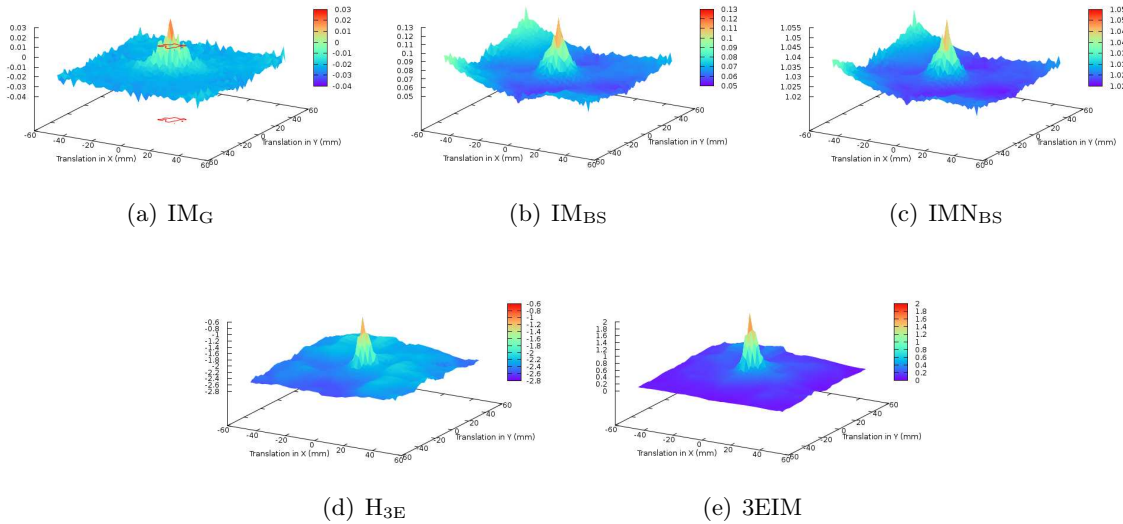


FIGURE 4.11 – Test de régularité des différentes mesures en présence de bruit additif gaussien d'écart-type 1, en utilisant 50 échantillons dans l'estimation des mesures. Ces figures représentent la valeur moyennée sur 100 tests des différentes mesures de similarité (en z) en fonction des translations (en x et y) considérées.

Dans ces tests, les résultats montrent une nette supériorité de nos mesures par rapport aux mesures classiques. Ceci est particulièrement visible quand le bruit additif devient important (figures 4.12 et 4.14). Même si la nature des images, qui présentent des zones noires dans les coins, expliquent sans doute en partie les mauvais résultats des mesures traditionnelles, on voit que leur robustesse à un bruit gaussien n'est pas très convaincante. Nous nous attendions à ce que nos mesures soient supérieures aux mesures classiques dans ce cas de figure. En effet, la distribution de probabilité gaussienne a tous ses cumulants d'ordre strictement supérieur à 2 qui sont égaux à 0. Or, nos mesures se basent sur les cumulants d'ordre supérieur à 2, et sont donc en théorie insensibles au bruit gaussien. Pour s'en convaincre, la figure 4.15 représente le résultat des simulations sur 3EIM, pour un bruit gaussien d'écart-type croissant et avec

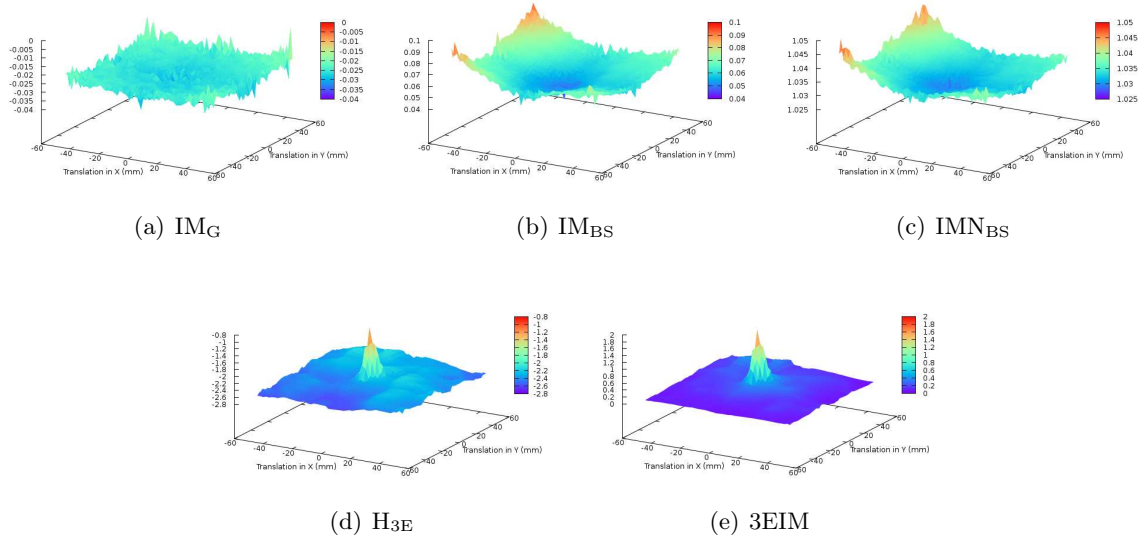


FIGURE 4.12 – Test de régularité des différentes mesures en présence de bruit additif gaussien d'écart-type 50, en utilisant 50 échantillons dans l'estimation des mesures. Ces figures représentent la valeur moyennée sur 100 tests des différentes mesures de similarité (en z) en fonction des translations (en x et y) considérées.

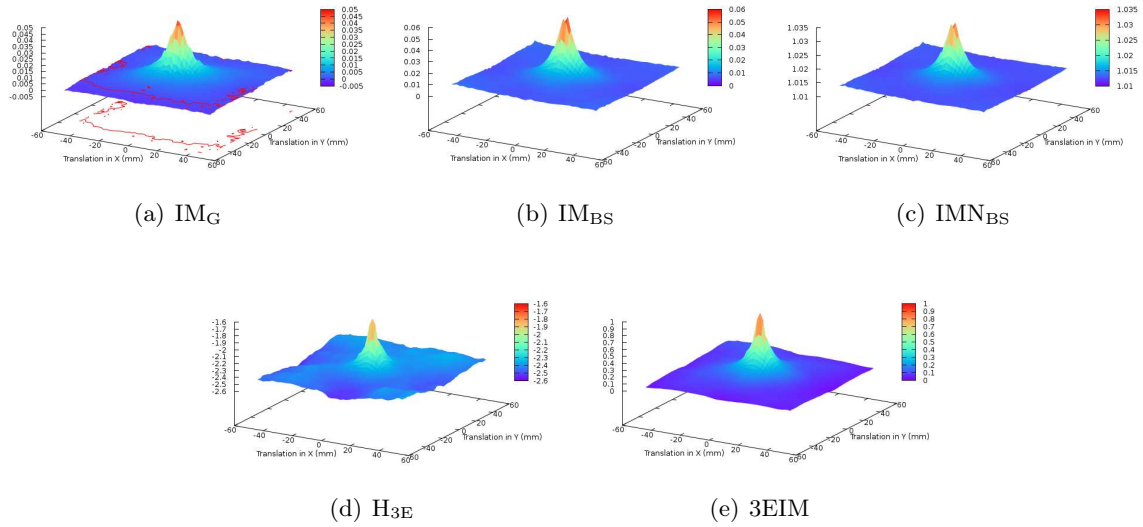


FIGURE 4.13 – Test de régularité des différentes mesures en présence de bruit additif gaussien d'écart-type 1, en utilisant 500 échantillons dans l'estimation des mesures. Ces figures représentent la valeur moyennée sur 100 tests des différentes mesures de similarité (en z) en fonction des translations (en x et y) considérées.

la prise en compte de 500 échantillons dans l'approximation de la mesure. On constate des résultats similaires quelle que soit la puissance du bruit. De plus, dans les 2 premières séries de tests, nous avons constaté que l'IM estimée par un fenêtrage de Parzen à noyau gaussien (IM_G)

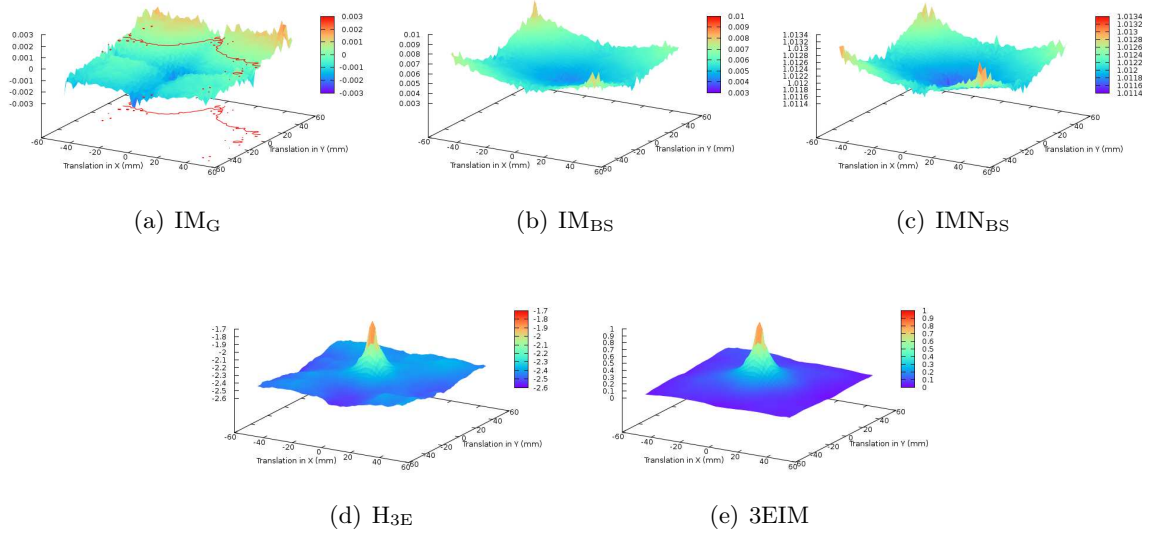


FIGURE 4.14 – Test de régularité des différentes mesures en présence de bruit additif gaussien d'écart-type 50, en utilisant 500 échantillons dans l'estimation des mesures. Ces figures représentent la valeur moyennée sur 100 tests des différentes mesures de similarité (en z) en fonction des translations (en x et y) considérées.

se comporte beaucoup moins bien que celle estimée par un noyau B-Spline (IM_{BS}), ceci étant particulièrement vrai quand le nombre d'échantillons est faible. Partant de cette constatation, nous avons décidé de ne conserver dans la dernière série de tests que les IMs estimées par un fenêtrage de Parzen à noyau B-Spline.

Cet exemple (figure 4.15) est intéressant car il montre la robustesse de notre mesure face au bruit additif gaussien, ce qui peut être un atout indéniable dans certains contextes applicatifs. Cependant, il nous a semblé nécessaire de faire de nouveaux tests qui prendraient en compte un autre type de bruit, qui n'avantagerait pas à priori nos mesures. Nous avons choisi dans ces simulations le bruit multiplicatif de speckle qui est connu pour dégrader les images échographiques. Une illustration de ce bruit est donné sur la figure 4.16.

A ce stade de l'évaluation, il nous reste donc 4 mesures : IM_{BS} et IMN_{BS} estimées par fenêtrage de Parzen à noyau B-Spline et l'entropie conjointe ainsi que l'IM basée sur un développement d'Edgeworth tronqué à l'ordre 3 (respectivement H_{3E} et 3EIM). Celles-ci ont été évaluées, en fonction du nombre d'échantillons pris en compte dans leur estimation, pour un bruit de speckle d'écart-type croissant de 0.05, 0.1 et 0.2, et en considérant encore une fois une translation de ± 50 mm en x et y . Nous ne donnerons ici qu'un exemple illustratif, car les 4 mesures se comportent globalement bien, ne présentant pas de minima locaux, quand le nombre d'échantillons est important (> 500). Cependant, quand ce nombre est faible (50), les résultats s'avèrent meilleurs pour nos mesures, comme nous pouvons le voir sur la figure 4.17. Ceci est un résultat intéressant car il montre qu'à priori, même en présence de bruit non-gaussien multiplicatif, notre mesure est régulière même quand elle est estimée à partir d'un faible nombre d'échantillons. Sur cette même figure 4.17, il est également intéressant de constater la grande différence entre IM_{BS} et IMN_{BS}. Ceci vient du fait que l'entropie conjointe moyenne (sur les 100 images) estimée est très proche de 1 quelle que soit la translation considérée, la majeure partie de l'information est alors portée par l'entropie marginale de l'image flottante.

Dans cette première partie des évaluations, nous avons comparé les mesures que nous

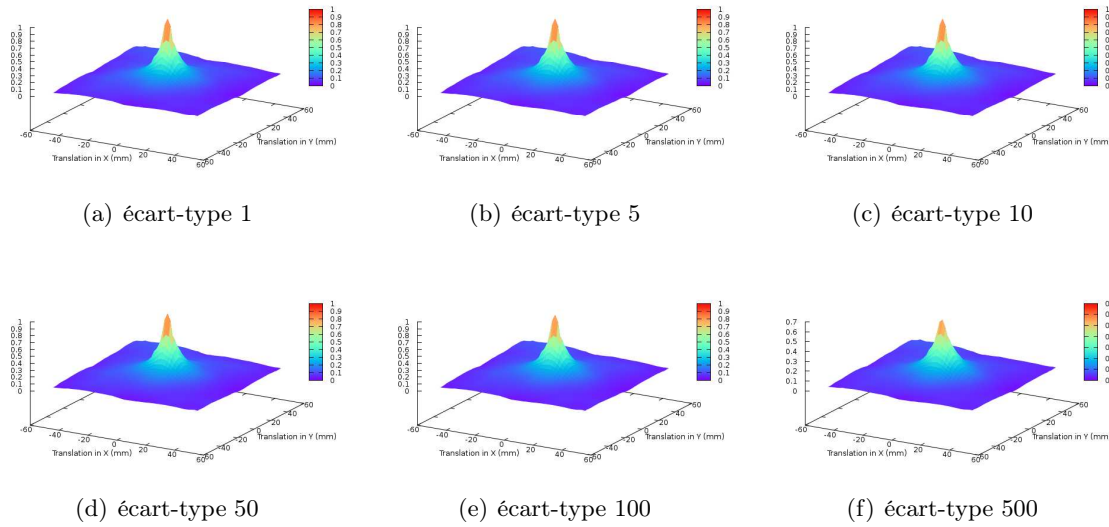


FIGURE 4.15 – Test de régularité de 3EIM en présence de bruit additif gaussien d'écart-type croissant, en utilisant 500 échantillons dans l'estimation des mesures. Ces figures représentent la valeur moyennée sur 100 tests des différentes mesures de similarité (en z) en fonction des translations (en x et y) considérées.

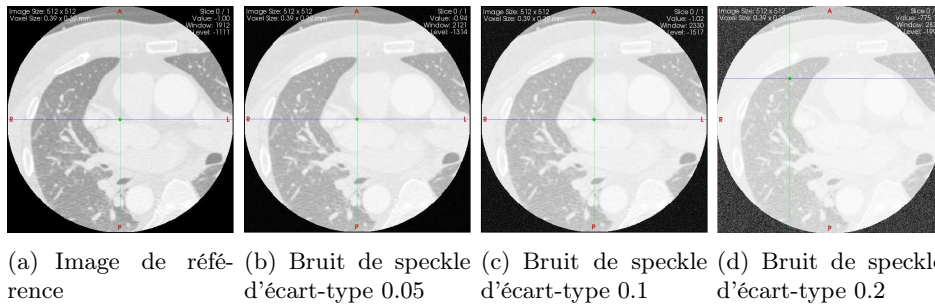


FIGURE 4.16 – Exemples d'images bruitées par un bruit multiplicatif de type speckle d'écart-type croissant.

avons développées aux mesures usuelles d'IMs sur des tests synthétiques, afin de regarder leur régularité dans l'espace des paramètres et lorsqu'elles sont soumises à différents types de bruits. Pour les mesures usuelles, les résultats montrent que l'IM estimée par un fenêtrage de Parzen à noyau B-Spline est plus régulière que celle estimée par noyau gaussien, surtout quand le nombre d'échantillons utilisés dans l'estimation est faible. L'IM estimée par histogramme est peu comparable, étant donné que d'après sa définition originale, elle prend en compte tous les pixels de l'image pour son estimation. L'évaluation de nos mesures a montré la singularité de celles estimées à partir d'un développement d'Edgeworth tronqué à l'ordre 4, et la bonne régularité de celles estimées à partir d'un développement d'Edgeworth tronqué à l'ordre 3 dans les différents scénarios envisagés. Plus particulièrement, les tests ont mis en évidence 3EIM et 3EIMN par rapport à l'entropie conjointe seule. Ces deux mesures donnant des résultats très similaires, nous ne gardons que la première dans la suite de notre étude. Enfin, si l'on compare 3EIM et IM_{BS} , nous obtenons sur ces tests synthétiques des résultats légèrement

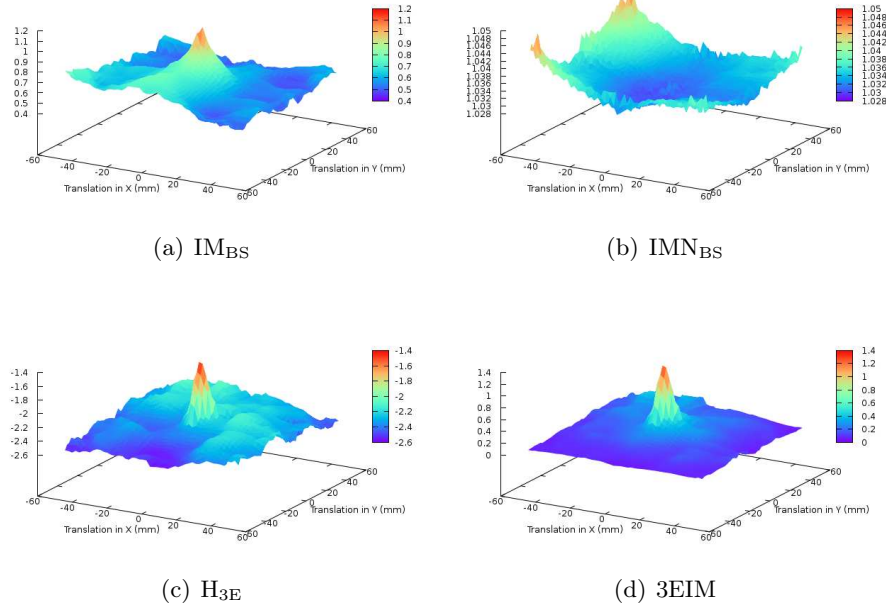


FIGURE 4.17 – Test de régularité des différentes mesures en présence de bruit multiplicatif de type speckle d'écart-type 0.2, en utilisant 50 échantillons dans l'estimation des mesures. Moyenne sur 100 tests.

supérieurs avec notre mesure, surtout lorsque le nombre d'échantillons est faible. L'étude de régularité ayant démontré les qualités de 3EIM en tant qu'approximation de l'IM, nous pouvons désormais l'intégrer dans un schéma de recalage global.

4.5 Evaluations des optimiseurs

Il nous a paru important de tester 3EIM avec différents optimiseurs et leurs jeux de paramètres associés. En effet, la mesure de similarité n'est qu'une des composantes d'une méthode de recalage, et même si elle semble régulière, il est important de l'insérer dans un schéma global. En nous basant sur les articles [Klein et al., 2007, Klein et al., 2009], nous avons tout d'abord sélectionné les optimiseurs qui nous paraissaient les plus adéquats pour résoudre notre problème. Le premier de ces articles présente une comparaison des différentes méthodes d'optimisation dans le cadre du recalage non-rigide d'images médicales utilisant l'IM comme mesure de similarité et les B-Splines comme modèle de déformation. Le second article, plus récent, présente une nouvelle méthode d'optimisation appelée descente de gradient stochastique à pas adaptatif, qui se veut plus efficace que toutes les méthodes proposées dans [Klein et al., 2007].

4.5.1 Présentation des paramètres des différents optimiseurs

Comme nous l'avons vu dans nos tests préliminaires sur les mesures de similarité, et notamment quand nous avons comparé le comportement de 3EIM par rapport aux IM classiques (les versions à noyau gaussien [Viola and Wells, 1997] et à noyau B-Splines [Thévenaz and Unser, 1997, Mattes et al., 2003]) sur des déformations simples, notre métrique ne se comporte pas exactement comme celles qui font référence dans le domaine. Aussi, il nous

apparaissait inopportun d'essayer d'utiliser directement les optimiseurs de la littérature avec les paramètres « optimaux » qui sont donnés. Nous avons donc testé nos métriques avec différents optimiseurs et en faisant varier leurs paramètres associés. Pour obtenir un résultat objectif de la qualité du recalage par rapport aux différents paramètres des optimiseurs, nous avons utilisé les données d'entraînement (1 patient) de la base de données RIRE [West et al., 1997] que nous présentons en détail dans l'annexe E. Cette base présente plusieurs avantages qui nous ont amené à l'utiliser. Tout d'abord, elle est reconnue par la communauté du recalage, puisqu'elle est souvent utilisée dans l'évaluation de nouvelles méthodes, comme nous l'avons dit au chapitre 2. De plus, elle présente des données multimodales scanner (CT), IRM (T1, T2 et R) et TEP et permet donc d'évaluer le recalage rigide multimodal. Enfin, et c'est le point le plus important, il est possible d'obtenir à partir de cette base de données une estimation quantitative de la qualité du recalage, par le biais des erreurs moyenne, médiane et maximale de recalage. La manière de calculer ces erreurs, qui sont des distances euclidiennes entre volumes d'intérêt, est présentée en détail dans l'annexe E.

Nous avons choisi de tester plusieurs optimiseurs donnés comme les plus performants dans la littérature, et qui ont été présentés au chapitre 1 :

- la descente de gradient (G), qui utilise une mise à jour du pas assez basique.
- le gradient conjugué non linéaire (GC), qui se sert d'une recherche linéaire inexacte pour calculer le pas à chaque itération, et d'une mise à jour de la direction de recherche prenant en compte le gradient mais aussi la direction de recherche à l'itération précédente.
- la descente de gradient stochastique à pas adaptatif, avec (GS_manu) ou sans (GS_auto) sélection manuelle des paramètres, dont l'apport principal est la sélection « intelligente » du pas à chaque itération.
- l'algorithme LBFGS (Limited memory Broyden-Fletcher-Goldfarb-Shanno), qui comme GC utilise une recherche linéaire inexacte pour calculer le pas, et approxime la matrice hessienne pour définir la direction de recherche.

Nous présentons dans ce qui suit les différents paramètres rentrant en jeu dans ces optimiseurs.

Paramètres liés au choix du pas

La détermination du pas est différente suivant l'optimiseur employé. Le pas de la descente de gradient est simplement un facteur de relaxation compris entre 0 et 1 qui guide la décroissance du pas initial. Quand l'optimiseur ne progresse plus, le pas initial est multiplié par ce facteur. Les 2 paramètres qui rentrent en jeu sont donc le facteur de relaxation (FR) et le pas initial (PI).

Le pas du gradient conjugué non linéaire et de l'algorithme LBFGS est déterminé itérativement par recherche linéaire inexacte décrite au chapitre 1. Cette procédure nécessite 4 paramètres :

- les constantes c_1 et c_2 issues des équations (1.23) et (1.24), qui dictent le comportement des conditions fortes de Wolfe. c_1 permet de régler le « taux » de décroissance que l'on désire, à chaque itération, sur la fonction de coût, tandis que c_2 dicte la manière dont on veut s'approcher d'un point stationnaire de cette fonction de coût.
- le pas initial de la recherche linéaire inexacte,
- le nombre maximum d'itérations pour cette recherche (ItLS).

Enfin, la descente de gradient stochastique à pas adaptatif utilise sa propre méthode également présentée au chapitre 1. Celle-ci prend en compte les paramètres a , A , α , f_{\min} , f_{\max} et ω et t_0 définis dans les équations (1.25) et (1.27). La force de cette méthode réside dans le fait qu'une grande partie de ces paramètres (a , f_{\min} , f_{\max} et ω) peuvent être estimés automatiquement, au début du recalage, à partir des images prises en compte, et plus précisément

à partir de leurs gradients et de leurs jacobiens. Mais ils peuvent également être donnés par l'utilisateur. Dans nos expérimentations, nous avons testé les 2 solutions.

Paramètres liés au choix de la direction de recherche

À l'heure actuelle, nous n'avons pas calculé de dérivée analytique de nos différentes mesures de similarités. C'est un travail nécessaire et qui fait partie des perspectives à court terme de ce travail. Or, les méthodes d'optimisation utilisées se basent toutes sur la direction du gradient de la fonction de coût. Nous avons donc utilisé une approximation $\tilde{\mathbf{g}}_k$ de la dérivée $\mathbf{g}(\boldsymbol{\mu}_k)$ calculée par différence finie. Cela nécessite de définir le scalaire c_k de l'équation (1.30). C'est le seul paramètre à fixer pour les méthodes de descente de gradient et de descente de gradient stochastique à pas adaptatif.

Pour la méthode de gradient conjugué non linéaire, il faut choisir la manière de calculer ζ_k défini dans l'équation (1.34), qui contrôle l'influence relative du gradient et de la direction de recherche de l'itération précédente pour le calcul de la direction de recherche à l'itération courante. Outre les deux expressions de ce scalaire données dans le chapitre 1 (équations (1.35) et (1.36)), d'autres expressions développées dans la littérature ([Dai, 2003], p.1318 et [Klein et al., 2007]) peuvent être utilisées.

Enfin l'algorithme du LBFGS nécessite le choix d'un paramètre qui va déterminer la « mémoire » de l'optimiseur, c'est à dire le nombre d'itérations passées utilisées dans la construction de l'approximation du hessien. Nous notons ce paramètre M .

Critères d'arrêt

Dans toutes ces méthodes, des critères d'arrêt doivent être également donnés, car dans le cas contraire ces optimiseurs itéreront indéfiniment. Les critères retenus sont :

- le nombre maximal d'itérations de l'algorithme (*ItMax*).
- le pas minimal (*PMin*). Au fur et à mesure que l'on se rapproche de l'optimum, le pas va décroître pour gagner en précision. Nous pouvons considérer qu'en dessous d'une certaine valeur de ce pas, on a atteint un optimum (et si possible l'optimum global).
- le gradient minimal (*GMin*). De la même manière, il est possible de considérer qu'en dessous d'une certaine valeur du gradient, l'optimum est atteint.
- pour les méthodes basées sur une recherche linéaire inexacte, on peut décider si l'on stoppe ou non l'algorithme lorsque les conditions fortes de Wolfe (équations (1.23) et (1.24)) ne sont pas satisfaites à une itération. Nous appelons ce critère Wolfe.

Schéma multirésolution

Finalement, nous avons voulu évaluer l'apport d'une procédure multirésolution dans la qualité de l'optimisation. Comme présentée dans le chapitre 1 (1.7), l'approche multirésolution peut intervenir à deux niveaux dans une procédure de recalage. On peut ainsi augmenter progressivement la complexité du modèle de déformation ou la complexité des données. Etant donné que cette première série de tests ne s'attache qu'à des déformations rigides, seule l'augmentation de la complexité des données nous intéresse ici. Aussi, avons nous choisi de comparer deux approches présentées en détail dans [Lester and Arridge, 1999] :

- une pyramide gaussienne avec sous-échantillonnage.
- une pyramide gaussienne sans sous-échantillonnage.

Un exemple de ces deux stratégies est donné figure 4.18.

Dans nos tests, nous avons donc considéré ces 2 types de pyramide, en utilisant 1 (approche monorésolution), 2, ou 3 résolutions. Quand l'approche multirésolution a été envisagée, nous avons utilisé un facteur 2 entre chaque niveau de la pyramide.

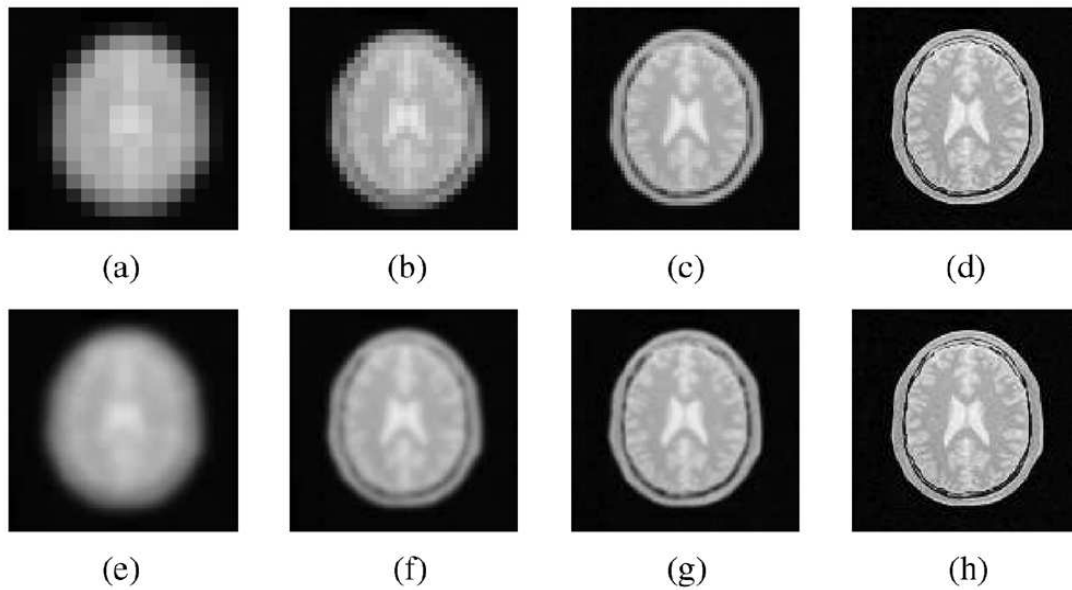


FIGURE 4.18 – Les 2 approches multirésolutions utilisées au cours des tests : en haut, une approche avec sous-échantillonnage. En bas, une approche sans sous-échantillonnage. D’après [Klein et al., 2010].

Comme nous l’avons constaté, ces différents optimiseurs possèdent de nombreux paramètres. Notre idée première était de déterminer les paramètres « optimaux » pour chaque optimiseur et pour recalcr « n’importe quel » type d’images. En effet, pour chaque patient de la base RIRE, 8 volumes 3D d’images (CT, TEP, IRM_T1, IRM_T2, IRM_PD, IRM_T1_rectifié, IRM_T2_rectifié, IRM_PD_rectifié) sont disponibles, et 12 recalages doivent donc être menés (6 recalages CT-IRM et 6 recalages TEP-IRM). En trouvant un jeu de paramètres permettant de recalcr correctement ces différentes modalités, on trouverait donc une méthode très générique pouvant s’adapter à beaucoup de problèmes de recalage. Une recherche exhaustive dans l’espace de ces paramètres est inenvisageable car elle engendrerait un nombre de recalages très important (et cela nécessiterait un optimiseur pour optimiser les paramètres de l’optimiseur, excepté si une solution analytique existe pour ces paramètres, ce qui consisterait à chercher les zéros de la dérivée de la fonction de coût par rapport à ces paramètres). Nous avons donc choisi empiriquement, avec l’aide des valeurs de référence données dans la littérature, différents jeux de paramètres pour les différents optimiseurs. Ceux-ci sont résumés dans le tableau 4.3*. Pour une question de présentation, ce tableau comme d’autres ont été placés à la suite de la bibliographie de ce chapitre. Nous les notons d’une * lorsque c’est le cas.

4.5.2 Evaluation

Dans toute cette partie, nous avons fixé le nombre d’échantillons utilisés par l’estimée la mesure de similarité à 5000. D’après les résultats de la section 4.4, ce chiffre donne une mesure de similarité régulière dans l’espace des transformations rigides. Cela permet de tester les optimiseurs sans que l’influence de la qualité de l’estimée de la mesure ne soit trop importante. Comme nous l’avons déjà dit, la base de données RIRE permet d’obtenir une évaluation quantitative de la qualité du recalage. Pour obtenir les résultats de recalage en ligne, nous avons écrit un script en langage Perl permettant de faire la requête automatiquement sur

le site, ce qui permet de recevoir le résultat automatiquement, et ainsi de réaliser un grand nombre de tests sans être obligé de se connecter à chaque fois au site internet. Pour le patient de la base de données d'entraînement, nous avons réalisé douze recalages 3D (6 recalages CT-IRM et 6 recalages TEP-IRM) pour chaque combinaison possible des paramètres décrits ci-dessus. Malheureusement, contrairement à ce que nous pouvions penser, nous n'avons pas trouvé, pour un optimiseur donné, un jeu de paramètres optimal pour tous les recalages différents. Aussi, pour évaluer la qualité des optimiseurs utilisés avec la mesure 3EIM, nous avons retenu trois critères :

- le pourcentage de recalage aboutis. Nous entendons ici les recalages qui ne divergent pas vers des valeurs aberrantes des paramètres de la transformation qui feraient « sortir » l'image flottante de son cadre.
- le pourcentage de recalages dont l'erreur moyenne était inférieure à 5mm. Cette valeur a été choisie pour donner une idée de la qualité des différents recalages. Nous ne l'avons pas choisie au hasard. Elle correspond approximativement aux performances données par plusieurs auteurs [Studholme et al., 1999, Pluim et al., 2001, Bardera et al., 2006] [Cahill et al., 2009] dans des articles évaluant différentes estimées de l'IM que nous avons citées au chapitre 3. De plus la taille de voxels des différents volumes de la base de données, entre les coupes, est de 4, 4 et 8mm pour respectivement le CT, l'IRM et la TEP. On peut donc grossièrement considérer qu'un recalage dont l'erreur moyenne est inférieure à 5mm a une précision voxelique en z . Enfin, comme expliqué dans l'annexe E, la précision du gold standard est de ± 0.6 mm pour les recalages CT/IRM et ± 2.4 mm pour le recalage TEP/IRM, donc choisir un seuil plus bas nous aurait amené aux limites proposées par l'approche, notamment pour le recalage TEP/IRM.
- les résultats moyens obtenus, pour chaque optimiseur, et pour tous les recalages considérés. Cela permet de se faire une idée précise de la sensibilité d'un optimiseur donné par rapport aux paramètres qui le décrivent.

Les résultats correspondant à ces trois critères, sont donnés respectivement dans les tables 4.4*, 4.5*, et les figures 4.19, 4.20, 4.21, 4.22 et 4.23.

4.5.3 Interprétation des résultats

Le première table 4.4* confirme la régularité de notre mesure (3EIM). En effet, quel que soit l'optimiseur utilisé (et le jeu de paramètres associés) et le couple d'images pris en compte, la méthode de recalage globale utilisant l'approximation de l'IM basée sur un développement d'Edgeworth tronqué à l'ordre 3 ne diverge que très rarement. Avec la descente de gradient standard, environ 10% des recalages divergent, mais cela est sans doute dû à la manière relativement « archaïque » de mettre à jour le pas de l'optimiseur. Les meilleures performances sont obtenues avec l'algorithme LBFGS et la descente de gradient stochastique à pas adaptatif, où la totalité des recalages convergent. Pour cette dernière méthode, la sélection manuelle des paramètres entraîne quelques recalages divergents.

Du second tableau 4.5*, où sont donnés le pourcentage des recalages pour lesquels l'erreur moyenne est inférieure à 5mm, nous pouvons tirer des conclusions très intéressantes pour la suite de notre travail. Tout d'abord, la méthode de descente de gradient stochastique à pas adaptatif dans laquelle les paramètres sont estimés automatiquement est de loin la plus efficace, puisque quels que soient les couples d'images considérés, environ 90% des recalages donnent cette erreur faible. Cela veut dire que cet algorithme, qui a déjà fait ses preuves en recalage d'images médicales [Klein et al., 2009], est adapté à notre mesure de similarité. L'autre leçon à tirer, c'est que notre mesure semble plus adaptée au recalage multimodal TEP-IRM que CT-IRM. En effet, lorsque l'on regarde les résultats correspondant aux algorithmes

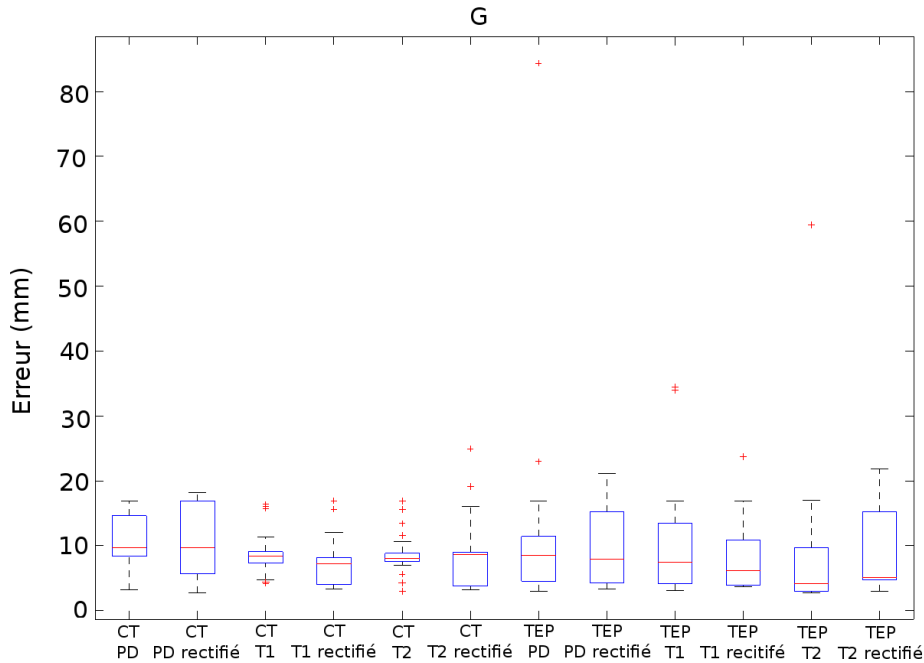


FIGURE 4.19 – Vue d’ensemble des recalages effectués, pour tous les paramètres, avec la descente de gradient (G). Ces boîtes à moustaches représentent les erreurs moyennes obtenues, pour chaque type de recalage, sur tous les tests effectués.

LBFSG et de gradient conjugué (GC), on observe une très grande différence entre les résultats pour les 2 couples d’imagerie pris en compte. Cela est sans doute dû à la différence dans la résolution spatiale entre les modalités d’imagerie. Une exception notoire est le recalage TEP-IRM_T2_rectifié qui donne de très mauvais scores avec ces 2 optimiseurs. Cela reste difficilement interprétable.

Les figures 4.19, 4.20, 4.21, 4.22 et 4.23 donnent une vue d’ensemble des recalages effectués pour tester les différents jeux de paramètres. Les résultats correspondant aux différents optimiseurs ne sont pas comparables entre eux, puisque suivant le nombre de paramètres testés, le nombre de recalages pour chaque optimiseur est différent. Avant d’analyser ces résultats, il est nécessaire d’expliquer clairement ce qu’ils représentent. Pour un optimiseur donné, nous avons fait varier ses paramètres comme indiqué dans le tableau 4.3, et nous avons fait, pour chaque combinaison de paramètres, les douze recalages (6 recalages CT-IRM et 6 recalages TEP-IRM) sur le patient d’entraînement de la base RIRE. Pour chaque recalage, nous avons obtenu une erreur moyenne de recalage. Ce sont ces erreurs moyennes, pour tous les recalages, qui sont données sur les figures 4.19, 4.20, 4.21, 4.22 et 4.23. Comme les paramètres que nous avons sélectionné se trouvent dans des intervalles de valeurs raisonnables, ces figures donnent donc une idée de la sensibilité de chaque méthode par rapport aux paramètres fournis. Notons enfin que ces figures ne contiennent pas les recalages non aboutis.

En ce qui concerne l’interprétation, les résultats pour la descente de gradient (figure 4.19) montrent une variance assez importante des résultats en fonction des paramètres choisis. De plus la valeur médiane est assez haute (entre 5 et 10 mm). Cela veut dire que cette méthode est sensible à l’initialisation des paramètres et peu efficace pour une utilisation avec notre mesure de similarité.

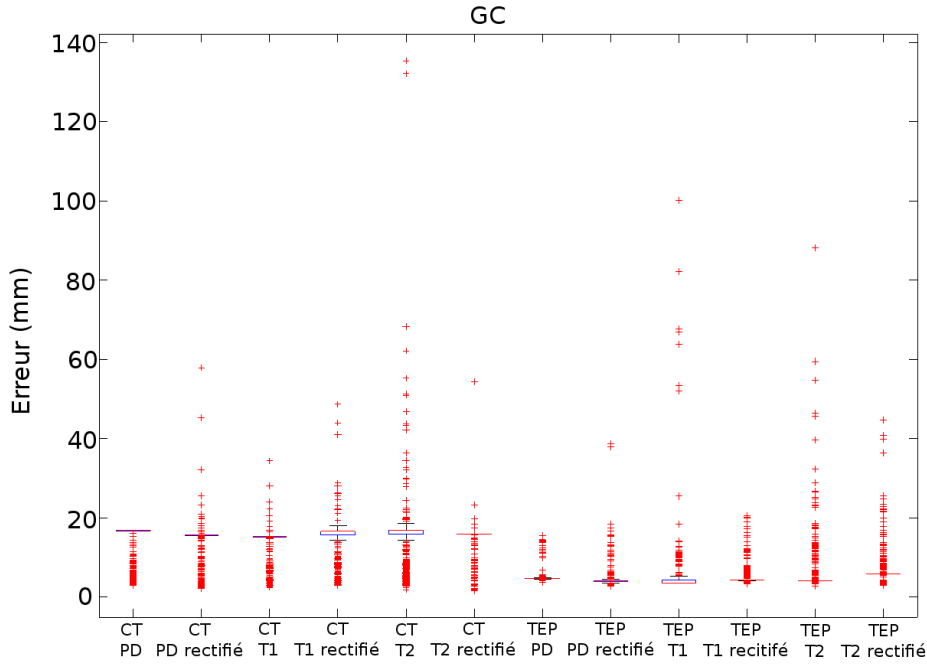


FIGURE 4.20 – Vue d’ensemble des recalages effectués, pour tous les paramètres, avec le gradient conjugué non linéaire (GC). Ces boîtes à moustaches représentent les erreurs moyennes obtenues, pour chaque type de recalage, sur tous les tests effectués.

La figure 4.20 est assez surprenante. Quels que soient les paramètres testés, le recalage avec le gradient conjugué non linéaire donne la même erreur moyenne, à quelques exceptions près (les valeurs aberrantes en rouge sur la figure). Cela veut dire qu’en association avec notre mesure de similarité, cet optimiseur est quasiment insensible aux paramètres initiaux. Par contre, les résultats moyens sont assez décevants pour le recalage CT-IRM (erreur moyenne d’environ 18mm), alors qu’ils sont assez satisfaisants pour le recalage TEP-IRM (erreur moyenne de 5mm). Les images TEP, qui sont très différentes des images CT et IRM, à la fois au niveau du contenu informatif, mais aussi de la résolution spatiale, expliquent sans doute ces résultats.

Les figures 4.21 et 4.22 donnent les résultats pour la descente de gradient stochastique à pas adaptatif, respectivement sans et avec sélection manuelle des paramètres. Quand on laisse l’algorithme déterminer lui-même les paramètres (à l’exception de A et t_0 qui doivent être fixés), les résultats sont toujours satisfaisants et les erreurs médianes (de l’ordre de 5mm) ont une variance faible (moins d’1 mm) en fonction des paramètres restés libres. Quand nous fixons nous même les paramètres, la variance augmente et le nombre de valeurs aberrantes également. Pour résumer, mieux vaut laisser l’optimiseur choisir lui-même les paramètres optimaux associés aux images à recalcr.

Enfin, l’algorithme LBFGS (4.23), tout comme GC, donne des résultats très différents suivant les modalités d’images considérées. Pour le recalage CT-IRM, il semble très sensible à l’initialisation et donne des résultats moyens, tandis qu’il n’est plus du tout sensible aux paramètres fixés et qu’il donne de bons résultats pour le recalage TEP-IRM (variance < 1 mm et valeur médiane d’environ 5 mm).

Si nous considérons les paramètres communs à tous les optimiseurs, nos expérimentations nous ont permis également de tirer quelques conclusions. Le critère d’arrêt qui dicte presque

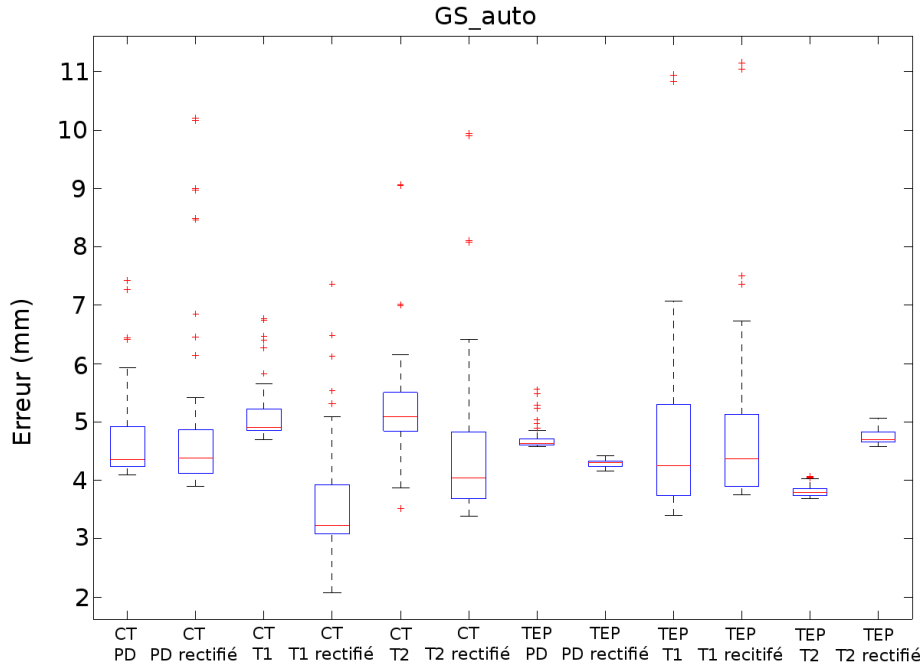


FIGURE 4.21 – Vue d’ensemble des recalages effectués, pour tous les paramètres, avec la descente de gradient stochastique à pas adaptatif, sans sélection manuelle des paramètres (GS_auto). Ces boîtes à moustaches représentent les erreurs moyennes obtenues, pour chaque type de recalage, sur tous les tests effectués.

toujours la fin du recalage est le nombre maximum d’itérations. Les quelques tests (non présentés ici) que nous avons réalisés en augmentant d’un facteur 10 ou 100 P_{min} et G_{min} , afin que ces critères d’arrêt soient mieux pris en compte dans le recalage, ont été décevants. Le recalage s’arrête alors trop tôt et trop loin de l’optimum. Pour les 2 méthodes utilisant la recherche linéaire inexacte, les conditions de Wolfe (quand le paramètre « Wolfe » est vrai) stoppent également souvent le recalage avant son terme. Nous avons également constaté qu’une valeur de 1 pour c_k donne les meilleurs résultats, et nous conservons donc cette valeur dans tout ce qui suit. Finalement, nous n’avons pas observé de différences très notables entre les 2 types de pyramides gaussiennes utilisées. Et plus surprenant, la qualité du résultat, dans ces tests, n’est pas dépendante du nombre de niveaux de pyramide utilisés.

A partir de tous ces résultats, nous avons donc décidé d’utiliser en association avec notre mesure de similarité l’algorithme de descente de gradient stochastique à pas adaptatif avec estimation automatique des paramètres initiaux, qui semble le plus efficace, mais également le plus robuste. Il faut garder en tête que notre but est de proposer une méthode de recalage globale pouvant s’adapter à plusieurs modalités d’imagerie. Cependant, pour des applications de recalage multimodal TEP-IRM, ou plus généralement lorsqu’une des modalités est la TEP, il faudra également considérer les algorithmes LBFGS et de gradient conjugué non linéaire, qui donnent de bons résultats et sont insensibles à leur initialisation dans ce cas.

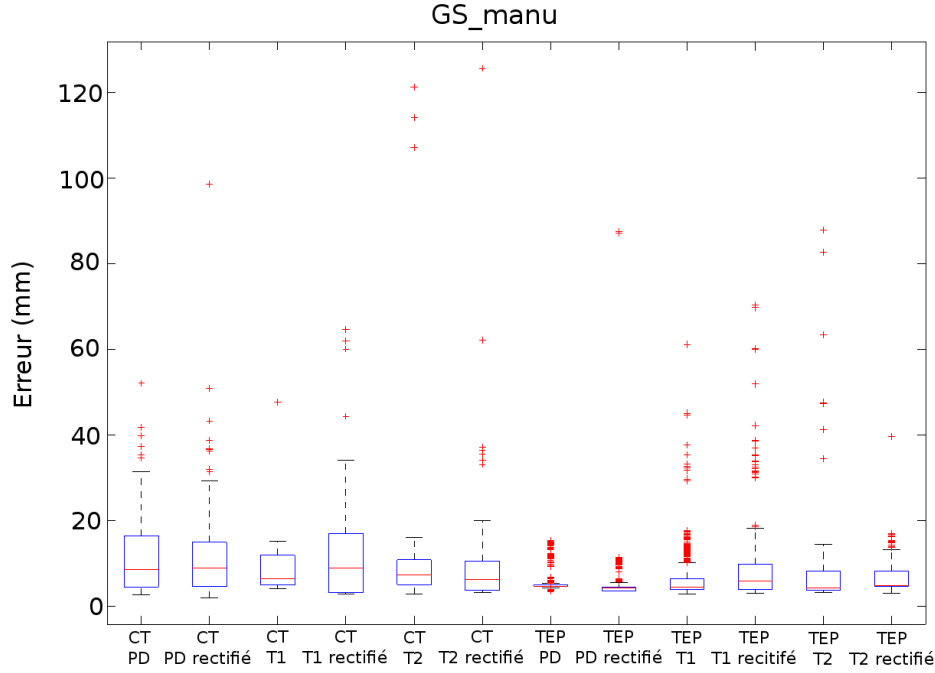


FIGURE 4.22 – Vue d'ensemble des recalages effectués, pour tous les paramètres, avec la descente de gradient stochastique à pas adaptatif, avec sélection manuelle des paramètres (GS_manu). Ces boîtes à moustaches représentent les erreurs moyennes obtenues, pour chaque type de recalage, sur tous les tests effectués.

4.6 Comportement de l'optimiseur en fonction du nombre d'échantillons

A ce point, nous disposons donc d'une mesure de similarité qui semble robuste, comme nous l'avons vu dans la section 4.4. Nous avons également réussi à l'insérer dans un schéma global de recalage rigide, ce qui nous a permis de choisir un optimiseur adéquat (GS_auto). Comme nous l'avons souvent dit dans la section 4.4, le nombre d'échantillons nécessaires pour obtenir une bonne estimation de la mesure de similarité est un paramètre important d'une méthode de recalage, puisqu'elle va grandement dicter le temps de calcul nécessaire pour une mesure donnée. Nous avons donc décidé d'étudier le comportement de l'optimiseur que nous avons sélectionné dans la section précédente (GS_auto) en fonction du nombre d'échantillons. Nous avons réalisé cette étude pour notre mesure 3EIM, mais également pour les mesures de référence IM_{BS} , IMN_{BS} et IM_G . Pour ce faire, nous avons suivi la même procédure que dans la section précédente (4.5), en faisant varier les paramètres de GS_auto comme décrit dans le tableau 4.3. Un nouveau paramètre vient s'ajouter à ceux-ci. Il s'agit du nombre d'échantillons, que nous avons pris égal à 50, 100, 500 et 1000. Nous présentons dans les figures 4.24, 4.25, 4.26 et 4.27 les résultats pour le cas du recalage CT-IRM_T1 par souci de concision, mais les résultats sont similaires pour les autres modalités d'imagerie considérées, et l'interprétation reste la même.

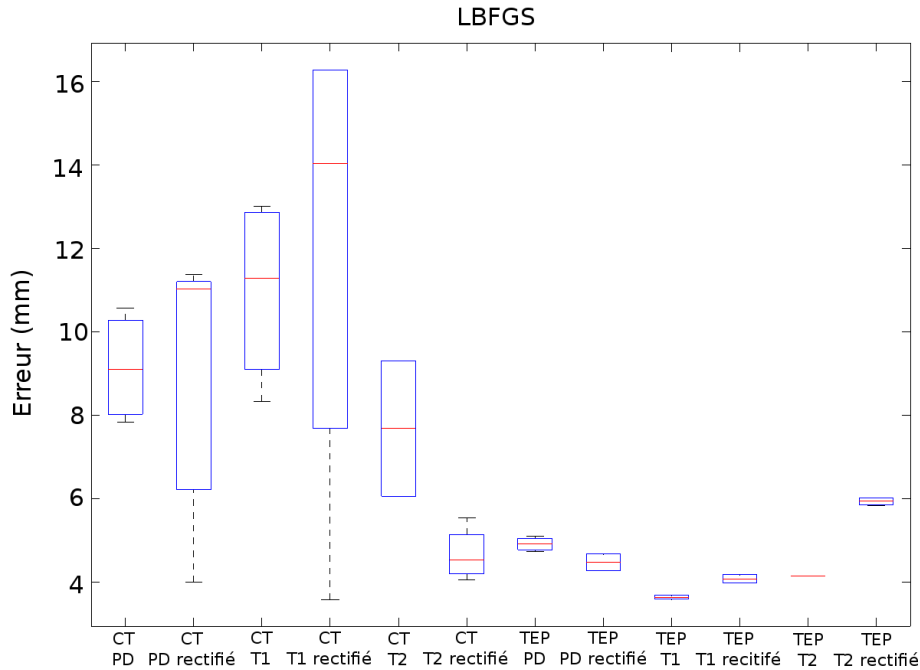


FIGURE 4.23 – Vue d’ensemble des recalages effectués, pour tous les paramètres, avec l’algorithme LBFGS. Ces boîtes à moustaches représentent les erreurs moyennes obtenues, pour chaque type de recalage, sur tous les tests effectués.

4.6.1 Interprétation

Commençons par aborder les temps de calcul. Le tableau 4.6* donne les temps de calcul moyen, pour 1000 itérations de la méthode de recalage rigide en 3 dimensions, en fonction de la mesure de similarité employée. Nous voyons tout d’abord que IM_G est très gourmande en temps de calcul. Plus grave, ce temps de calcul croît très fortement lorsque le nombre d’échantillons augmente. Cela s’explique par la construction de l’histogramme conjoint et du noyau gaussien utilisé, qui considère un voisinage important pour lisser l’histogramme. Ensuite, nous constatons que notre mesure est moins rapide que les IM s estimées par fenêtrage de Parzen à noyau B-Spline, le temps de calcul de ces 3 mesures augmentant tout de même linéairement en fonction du nombre d’échantillons. L’implémentation joue assurément un rôle dans ces différences. En effet, les méthodes IM_{BS} et IMN_{BS} étant quasiment similaires (à l’exception d’une soustraction qui est remplacée par une division dans le calcul), on ne devrait pas avoir une telle différence dans les temps d’exécution liés à ces 2 mesures. Les fonctions utilisées dans ces tests sont celles de la librairie *elastiX*, et les différences entre ces 2 mesures viennent donc sans doute des différences d’implémentation de ces 2 méthodes. Pour ce qui est de notre mesure, nous l’avons initialement implémentée dans la structure de la librairie *ITK*, et l’avons seulement adaptée pour l’utiliser avec *elastiX*. Cependant, elle ne tire pas partie de toutes les capacités de cette seconde librairie, et notamment de la classe *AdvancedImageToImageMetric* qui permet d’améliorer le rendement des mesures de similarité. Cela demande un travail important de réimplémentation qui sera nécessaire à court terme, et nous pensons que cela permettra d’augmenter significativement les performances de notre mesure.

Ceci étant dit, étudions la figure 4.27. Quand le nombre d’échantillons est faible (50, 100),

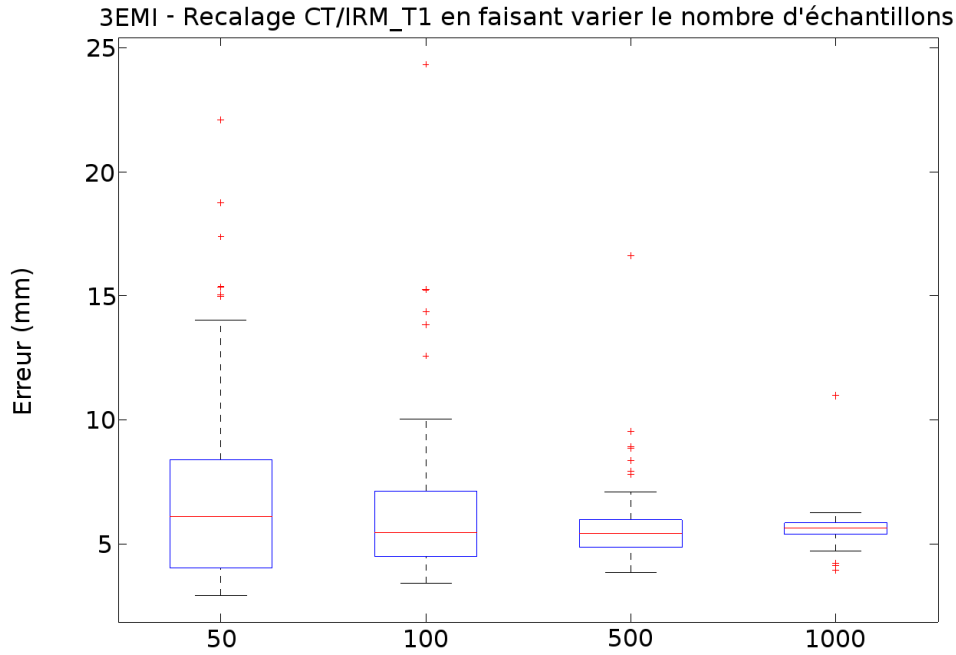


FIGURE 4.24 – Qualité du recalage CT-IRM_T1 en fonction des paramètres de l'optimiseur et du nombre d'échantillons utilisés pour l'estimation de 3EIM.

IM_G est complètement insensible aux paramètres de l'optimiseur et donne presque toujours la même erreur moyenne de recalage. Si nous nous référons à la section 4.4, nous avons vu que cette mesure n'était pas régulière pour un nombre si faible d'échantillons. Cela laisse supposer que dans ces expérimentations, l'optimiseur est rapidement bloqué dans un maximum local et ne converge donc pas vers la solution globale. L'erreur médiane trouvée dans ces situations (environ 17,5 mm) montre que le recalage est défaillant. Quand le nombre d'échantillons augmente, les résultats s'améliorent mais les recalages sont globalement meilleurs avec 500 plutôt que 1000 échantillons, ce qui est contraire à ce que laisserait supposer la théorie. Nous aurions sans doute dû augmenter encore le nombre d'échantillons pour voir l'influence sur les résultats, mais le coût calculatoire de cette mesure au delà de 1000 échantillons est irraisonnable pour une utilisation en recalage.

Comparons maintenant notre mesure 3EIM aux mesures de référence qui semblaient les plus régulières lors de l'évaluation des mesures (4.4), c'est à dire IM_{BS} et IMN_{BS} . Nous notons ici une différence remarquable. Pour les 2 mesures de référence (figures 4.25 et 4.26), l'augmentation du nombre d'échantillons semble améliorer fortement la qualité du recalage, puisque l'erreur moyenne médiane passe d'environ 12 mm à 7 mm. Pour notre mesure 4.24, l'augmentation du nombre d'échantillons n'améliore pas cette erreur moyenne d'environ 6 mm atteinte dès 50 échantillons. Par contre, la variance des erreurs médianes est réduite drastiquement, ce qui n'est pas le cas pour les mesures de référence. Cela nous amène à deux conclusions concernant notre mesure de similarité. Tout d'abord, elle est globalement suffisamment bien estimée avec un nombre d'échantillons réduit pour mener à bien un recalage rigide. C'est un point essentiel puisque cela veut dire que les temps de calcul s'en verront d'autant diminués. Schématiquement, la qualité du recalage avec notre mesure et 50 échantillons est comparable à celle de IM_{BS} et IMN_{BS} avec 1000 échantillons. Si l'on reprend le tableau 4.6*, cela veut

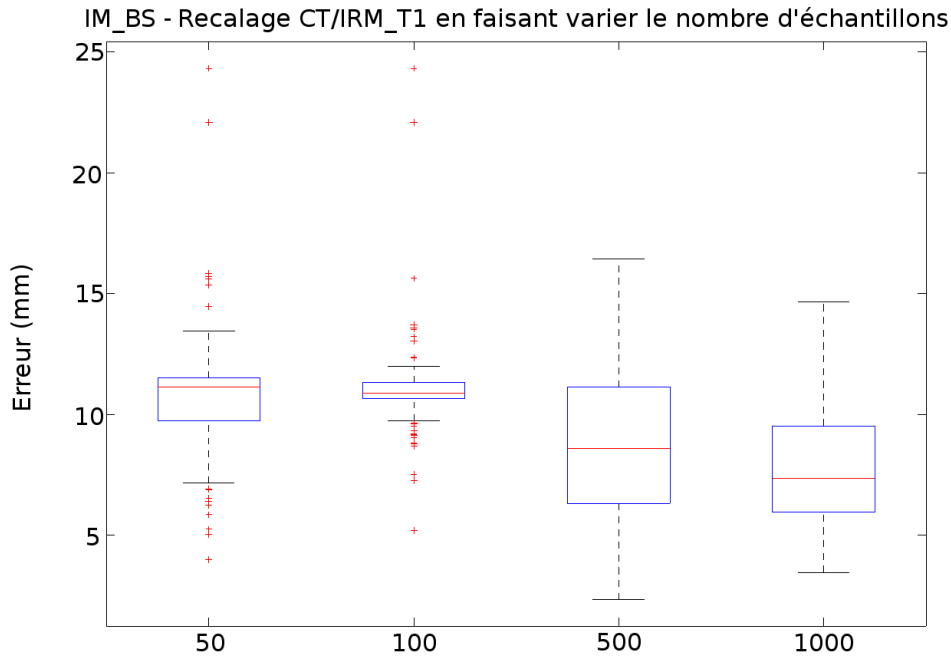


FIGURE 4.25 – Qualité du recalage CT-IRM_T1 en fonction des paramètres de l'optimiseur et du nombre d'échantillons utilisés pour l'estimation de IM_{BS} .

dire que l'on obtient finalement des temps de calcul meilleurs que ceux des méthodes de référence. Le deuxième enseignement est qu'en augmentant le nombre d'échantillons, notre mesure devient moins sensible aux paramètres de l'optimiseur. Parmi ceux-ci, on compte le nombre maximum d'itérations et le fait d'utiliser ou non un schéma multirésolution. On peut donc penser que notre mesure nécessite moins d'itérations que ses consœurs pour converger vers une solution acceptable, ou qu'elle est moins sensible à l'utilisation d'un schéma multirésolution. Ces hypothèses, si elles sont vérifiées, permettent encore une fois de gagner en temps de calcul. Elles seront développées dans la section suivante. Pour le recalage rigide et ses 6 paramètres en 3 dimensions, ce temps de calcul n'est évidemment pas une priorité avec nos mesures, car un nombre très restreint d'échantillons ou d'itérations permet de gagner au mieux quelques secondes. Par contre, quand nous considérons des déformations non-rigides pouvant prendre en compte plusieurs milliers de paramètres, cela devient primordial.

Dans cette partie, nous avons comparé le comportement de notre mesure et de celles de référence en fonction du nombre d'échantillons utilisés pour leur estimation. Pour ce faire, nous avons utilisé la base de données d'entraînement de RIRE, qui est composée d'un patient et de 8 volumes d'images multimodales. Pour valider définitivement l'utilisation de cette mesure en recalage d'images médicales, il nous reste cependant 2 tâches importantes. Il nous faut évaluer cette mesure face à des déformations non-rigides, qui sont bien plus complexes que les déformations considérées jusqu'à présent. Nous y viendrons dans la section 4.8.1. Mais tout d'abord, nous devons évaluer notre mesure pour le recalage rigide sur un jeu de données plus important. Pour ce faire, nous avons encore une fois utilisé la base RIRE, mais nous nous sommes servis cette fois-ci des données de tests, qui comptent dix-huit patients. C'est ce que nous présentons dans ce qui suit.

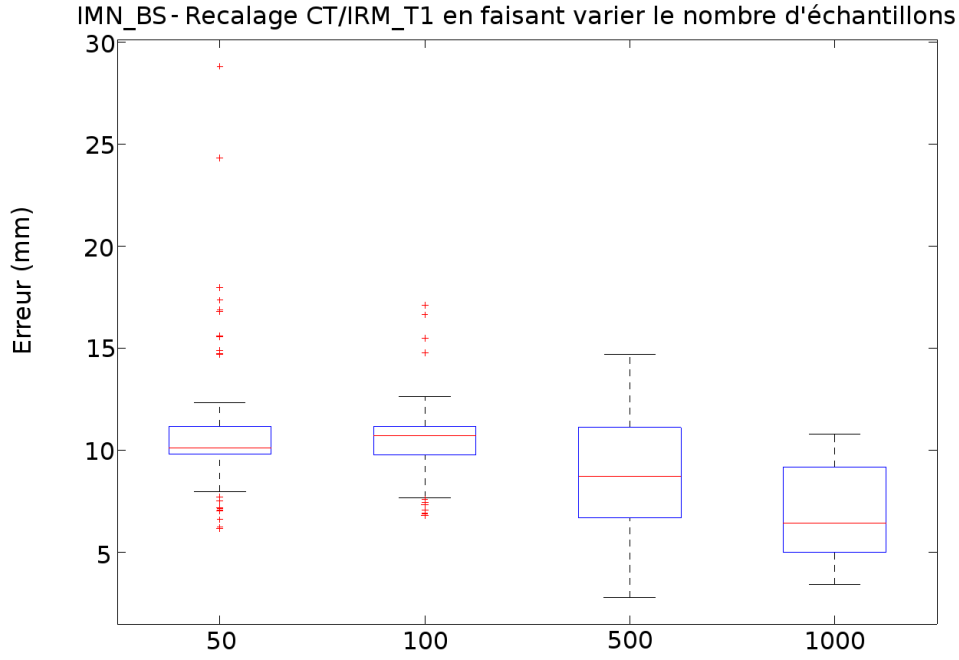


FIGURE 4.26 – Qualité du recalage CT-IRM_T1 en fonction des paramètres de l’optimiseur et du nombre d’échantillons utilisés pour l’estimation de IMN_{BS} .

4.7 Evaluation du recalage rigide

Dans cette partie, nous avons fixé les derniers paramètres libres de l’optimiseur, en posant $A = 20$ et $t_0 = 0$, qui sont les valeurs recommandées pour cet optimiseur [Klein et al., 2009]. Les derniers paramètres de la méthode de recalage sont alors :

- le nombre d’itérations. Nous le faisons varier dans les tests préliminaires et le fixons ensuite.
- le schéma multirésolution. Nous verrons qu’il a une grande importance pour les méthodes de recalage utilisant les mesures de similarité de référence.
- le nombre d’échantillons. Nous l’avons fixé à 1000 pour avoir une estimation assez précise des mesures de similarité, mais surtout pour pouvoir comparer les différentes méthodes entre elles, puisque nous avons vu que IM_G est difficilement utilisable en pratique avec un nombre supérieur d’échantillons

Nous avons considéré les 9 premiers patients de la base donnée par le tableau 4.7*. Un exemple des images de cette base est représenté sur la figure 4.28. La base de données ne contient pas les 8 modalités d’imagerie pour chaque patient, et nous avons en plus écarté les volumes CT du patient 004 et TEP du patient 005, qui donnaient des résultats aberrants dans les tests préliminaires, quelle que soit la mesure de similarité choisie. Ceci est dû à la mauvaise qualité de ces 2 volumes. Encore une fois, notre mesure 3EIM est comparée aux mesures de référence IM_G , IM_{BS} et IMN_{BS} . Nous ne donnons ici que les résultats les plus parlants, mais la totalité des tests réalisés est disponible en ligne, à l’adresse <http://www.insight-journal.org/rire/>, sous l’onglet Results (et à mon nom). Nous donnons dans la dernière colonne de chaque tableau de résultat la référence (sous la forme Es-saiX_X) vers les résultats disponibles en ligne.

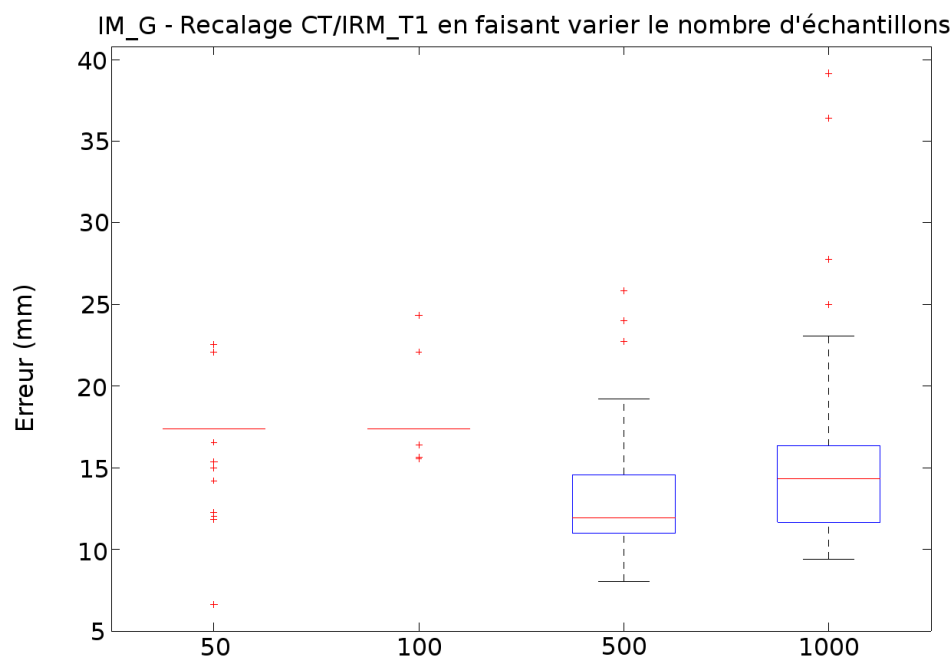


FIGURE 4.27 – Qualité du recalage CT-IRM_T1 en fonction des paramètres de l'optimiseur et du nombre d'échantillons utilisés pour l'estimation de IM_G .

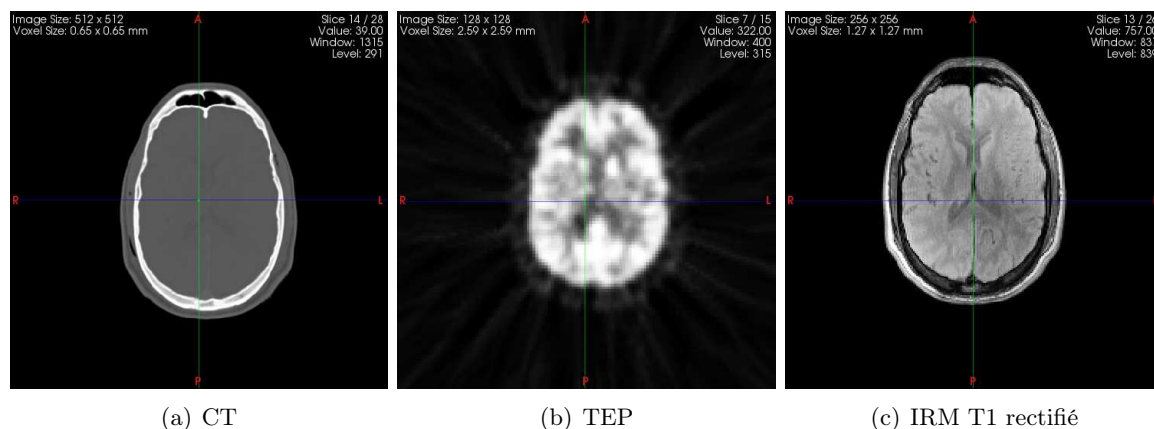


FIGURE 4.28 – Exemple de coupes d'images de la base de données RIRE.

4.7.1 Schéma monorésolution

Nous avons commencé par tester l'influence du nombre d'itérations, en utilisant les images en pleine résolution (1 seul niveau de pyramide). Nous avons choisi 200 et 1000 itérations, la première valeur étant recommandée lorsque l'on désire un recalage rapide, la deuxième quand une meilleure précision est nécessaire. Les différentes mesures de similarité sont estimées avec 1000 échantillons, pour une comparaison objective. Les résultats sont donnés dans le tableau 4.8*, et les meilleurs résultats pour chaque couple de modalités sont donnés en **gras**. Ce tableau donne, pour chaque couple de modalité d'imagerie, l'erreur moyenne de recalage

sur les 9 premiers patients de la base, en fonction de la mesure de similarité choisie et du nombre d'itérations. Le premier constat, c'est que le nombre d'itérations, à condition d'être choisi de manière « raisonnable », n'influence pas le résultat final. Les résultats sont même souvent meilleurs, pour une mesure de similarité donnée, pour 200 que pour 1000 itérations. Mais ce qui est surtout très intéressant, c'est que notre mesure est plus efficace que celles de référence, puisqu'elle obtient le meilleur score quelles que soient les modalités considérées. Il semble en effet que les mesures de référence ne permettent pas un recalage correct en ne considérant qu'un schéma monorésolution. Les deux hypothèses que nous émettions dans la section précédente (4.6) se vérifient : notre mesure (comme celles de référence) permet à l'algorithme de converger en un nombre d'itérations relativement faible, et elle ne nécessite pas à priori, contrairement à celles de référence, de schéma multirésolution pour le recalage rigide. Des exemples de résultats obtenus avec notre mesure sont donnés sur la figure 4.29. Pour se faire une idée visuelle des résultats, l'un de moyens les plus efficaces est d'utiliser un damier composé de parties des images de référence et flottante. Ainsi, les figures 4.29(a) et 4.29(b) montrent deux coupes comprenant des parties d'images CT et IRM, respectivement avant et après le recalage. La figure 4.29(c), quant à elle, donne tout son sens à ce travail de recalage : il permet de superposer les informations complémentaires provenant de plusieurs modalités d'imagerie différentes. Enfin, la figure 4.29(d) donne l'apport du CT (en bleu) et de l'IRM (en rouge) dans l'image composite. Des résultats similaires sont proposés pour le recalage TEP/IRM sur la figure 4.30.

Après ces premiers tests, nous avons voulu quantifier l'apport d'une approche hiérarchique sur la qualité du recalage, car les résultats des mesures de référence nous ont semblés étonnamment médiocres, par rapport aux références données dans la littérature [Studholme et al., 1999, Pluim et al., 2001, Bardera et al., 2006], qui font elles usage de la multirésolution.

4.7.2 Schéma multirésolution

Pour commencer, nous avons tenté de déterminer le nombre de niveaux optimal de la pyramide multirésolution. Nous avons pour cela effectué les recalages CT/IRM du patient 001, en faisant varier le nombre de niveaux de la pyramide d'images de 1 à 5, en utilisant notre mesure de similarité estimée à partir de 5000 échantillons (pour que les résultats ne soient pas trop dépendants de la qualité d'estimation de la mesure), et un nombre maximal d'itérations fixé à 200. Les résultats sont donnés dans le tableau 4.9*, avec en **gras** les meilleurs résultats pour un couple d'images donné. Nous constatons que la pyramide à 5 niveaux donne le meilleur résultat dans 3 cas sur 6, et nous avons donc décidé de retenir 5 niveaux de pyramide pour lancer les tests sur toutes les mesures de similarité et les 9 patients de la base. Mais ce que nous pouvons aussi constater, c'est que les résultats, en utilisant notre mesure de similarité, ne s'améliorent pas énormément en fonction des différentes approches.

Nous donnons dans le tableau 4.10* les résultats obtenus, pour les différentes mesures, sur les 9 patients de la base, avec une pyramide d'images à 5 niveaux et 200 itérations par niveau. Les résultats des mesures IM_{BS} , IMN_{BS} sont en nette amélioration, tandis que ceux de IM_G sont vraiment très mauvais, et qui plus est dans un temps de calcul beaucoup plus important. Nous pensons que l'utilisation d'un trop grand nombre de résolutions, avec cette mesure, tend à faire diverger l'algorithme pour les plus basses résolutions. Notre mesure donne encore le meilleur résultat dans la majorité des cas. Le recalage TEP/IRM, notamment, qui fait intervenir des modalités très différentes, semble beaucoup plus efficace avec l'utilisation de 3EIM plutôt qu'avec les mesures de référence. Les résultats pour IM_{BS} et IMN_{BS} sont globalement en adéquation avec les résultats de la littérature. De plus, IMN_{BS} donne de meilleurs résultats que IM_{BS} , comme prévu dans la théorie.

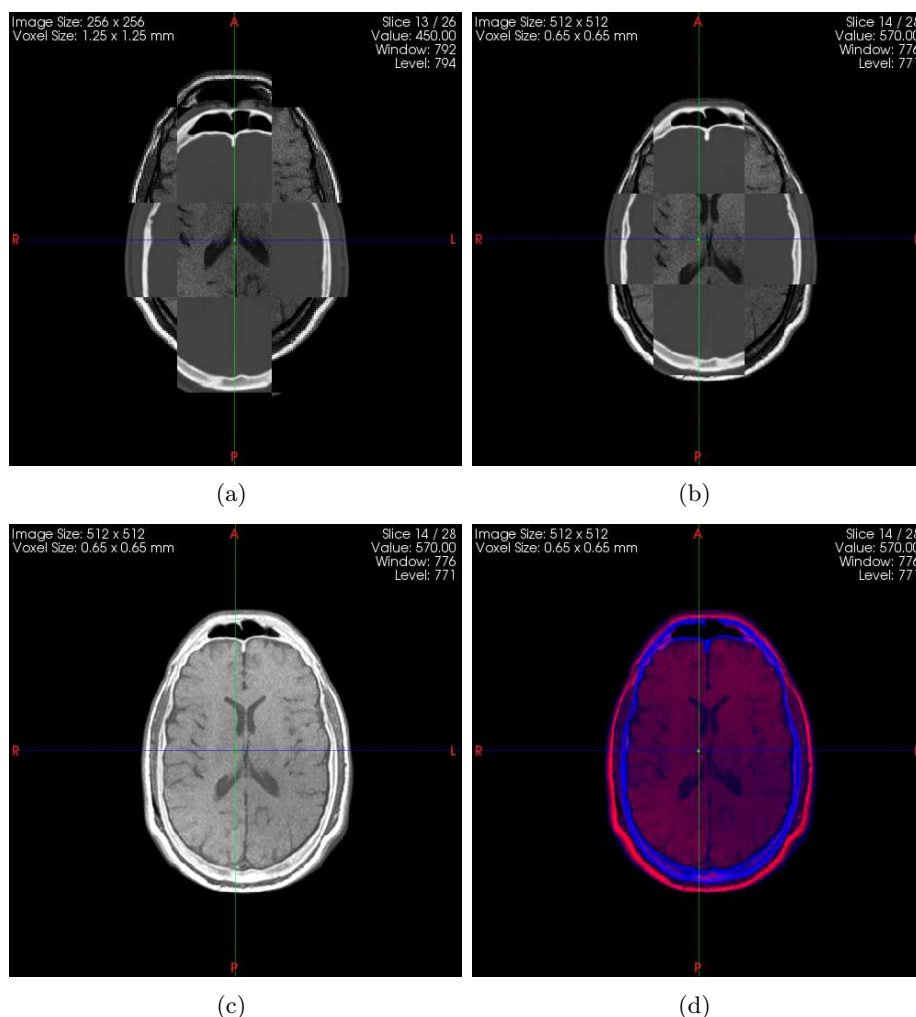


FIGURE 4.29 – Exemple de recalage CT/IRM effectué avec 3EIM. La figure 4.29(a) est une coupe, sous forme de damier, des 2 images avant le recalage. La figure 4.29(b) donne la même figure après le recalage. Enfin, les figures 4.29(c) et 4.29(d) sont des images composites des 2 modalités d'imagerie avec en bleu le CT et en rouge l'IRM sur la figure 4.29(d).

Dans cette partie, nous avons vu que notre approximation de l'IM permet un recalage rigide très efficace, et même globalement supérieur à celui obtenu avec les mesures de référence. Nous avons notamment constaté qu'elle autorise un recalage TEP/IRM de qualité. En pratique, le recalage rigide est souvent une étape préliminaire au recalage non-rigide (comme nous l'avons vu dans le chapitre 1, 1.5), car cela permet de mettre grossièrement en correspondance les images. Tous ces tests nous ont donc permis d'affiner ce prétraitement indispensable et de tirer quelques conclusions. Tout d'abord, les temps de calcul restent relativement faibles en recalage rigide (par rapport au non-rigide), car le nombre de paramètres à optimiser est beaucoup plus faible. Quand le recalage rigide est utilisé comme prétraitement, nous pouvons donc utiliser un nombre important d'échantillons, ainsi qu'un schéma multirésolution car ces 2 paramètres améliorent sensiblement les résultats. Dans la suite, nous présentons nos évaluations pour le recalage non-rigide, en utilisant ce prétraitement.

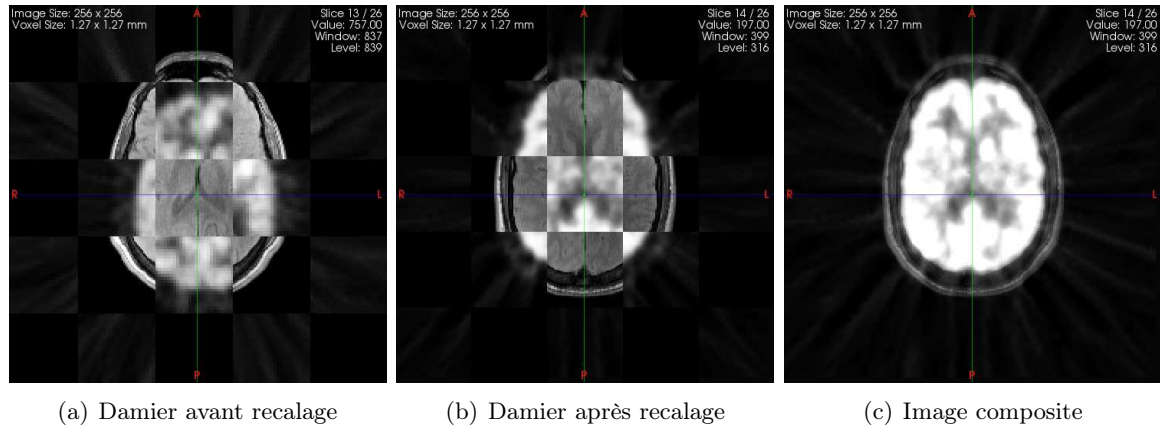


FIGURE 4.30 – Exemple de recalage TEP/IRM effectué avec 3EIM. Le principe est le même que sur la figure 4.29.

4.8 Evaluation du recalage non-rigide

L'évaluation qualitative du recalage non-rigide est une tâche relativement compliquée, et c'est un sujet de recherche très actif à l'heure actuelle [Škerl et al., 2006, Murphy et al., 2008, Kabus et al., 2009]. Nous n'avons pas trouvé de manière simple d'évaluer notre mesure pour des tâches de recalage multimodal non-rigide, car l'évaluation est toujours sujette à caution. Nous nous sommes donc concentrés sur le recalage monomodal CT/CT, et nous avons choisi 2 méthodes d'évaluation. La première se base sur le coefficient de Dice, qui mesure le recouvrement entre 2 segmentations recalées, et sur lequel nous reviendrons dans ce qui suit. Nous avons pour ce faire utilisé une base de données CT inter-patients du laboratoire. Mais n'étant pas totalement satisfaits de l'évaluation quantitative proposée par ce coefficient, comme nous l'expliquerons, nous nous sommes alors tournés vers la base de données CT intra-patients EMPIRE, qui a été utilisée lors du challenge MICCAI2010, et qui évalue la qualité du recalage à partir de 4 scores différents. Nous en parlerons dans un second temps. Dans cette partie, nous avons seulement utilisé notre mesure de similarité 3EIM. Pour la base de données du laboratoire, des études menées par d'autres doctorants [Dréan et al., 2011] permettent de comparer nos résultats. Pour la base EMPIRE, les résultats du challenge MICCAI2010 sont en ligne (<http://empire10.isi.uu.nl/mainResults.php>) et permettent de comparer notre méthode aux méthodes de pointe à l'heure actuelle.

4.8.1 Base de données du laboratoire

Nous commençons par présenter le contexte clinique, puis nous situons notre travail de recalage non-rigide. Enfin, nous donnerons les résultats obtenus.

Contexte clinique

Dans le traitement du cancer de la prostate par radiothérapie, une segmentation précise de la prostate et des organes à risque qui l'entourent sur les images scanner (CT) de planification est une tâche essentielle pour le planning thérapeutique. En effet, à partir des images prises avant le traitement, le médecin va décider de la quantité et de la localisation du dit traitement, le but étant d'irradier la tumeur située dans la prostate et dans les vésicules séminales, en essayant de préserver au maximum les tissus et organes environnants (vessie, rectum, têtes fémorales).

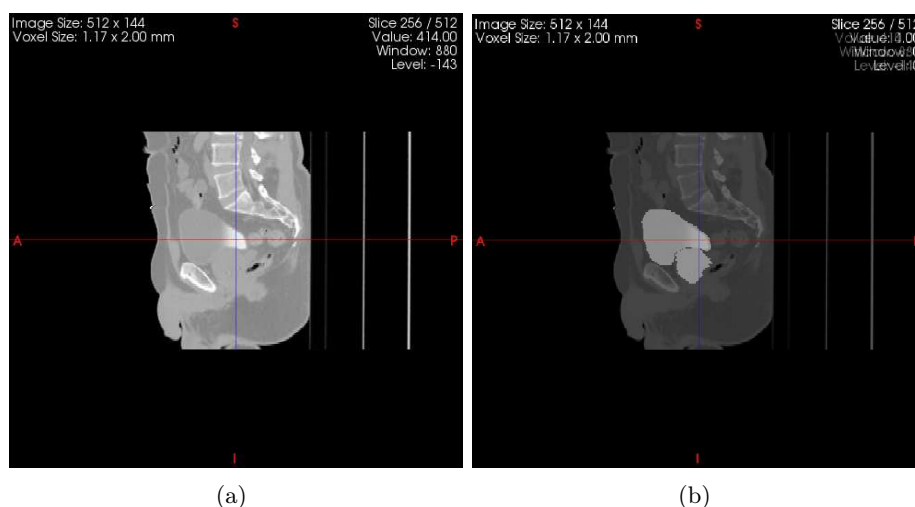


FIGURE 4.31 – Illustration de la difficulté d'une segmentation automatique. 4.31(a) : image originale. 4.31(b) : image sur laquelle ont été superposées les segmentations expertes de la vessie (en haut) et de la prostate (en bas). Si l'on regarde sur l'image originale, on se rend compte du faible contraste entre ces 2 organes.

Il est donc nécessaire de segmenter ces différents organes sur les CT de planification. Pour l'instant, c'est en général les médecins experts qui réalisent manuellement ces segmentations, ce qui prend beaucoup de temps et peut amener des erreurs dues à la variabilité inter-experts. La segmentation automatique est très délicate car le contraste entre les organes d'intérêt, comme illustré sur la figure 4.31, est très faible.

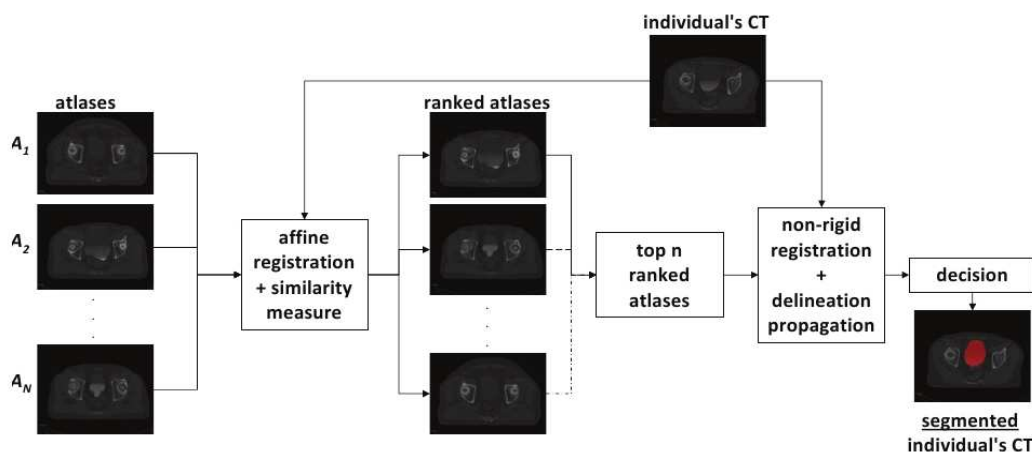


FIGURE 4.32 – Approche multi-atlas proposée au laboratoire. Les atlas sont classés en fonction de leur similarité par rapport à l'image étudiée. Puis les labels des n atlas les mieux classés sont propagés sur l'image étudiée, et une règle de décision permet d'obtenir les segmentations des organes. D'après [Acosta et al., 2011].

Une approche multi-atlas, développée au laboratoire, essaie de se prémunir à terme de ces segmentations délicates. Le principe de cette approche est illustré sur la figure 4.32. Nous disposons de 24 Atlas (1 atlas correspond au volume CT d'un patient, et des segmentations

manuelles expertes des organes d'intérêt, appelées labels) notés A_1, \dots, A_N , et du volume CT du patient (individual's CT sur la figure) noté I dans ce qui suit. L'approche se décompose en 3 étapes :

- i) Tout d'abord, les atlas recalés affinement sur I sont classés en fonction de leur similarité par rapport à I .
- ii) Puis les labels des n meilleurs atlas, $n \in N$, c'est à dire des atlas les plus similaires à I , sont propagés sur I en utilisant un recalage non-rigide. Expliquons cette procédure pour un atlas A_i donné : on procède au recalage non-rigide entre l'atlas A_i et I . Cela permet d'obtenir un champ dense de déformation non-rigide. A l'aide de ce champ, on déforme les labels de l'atlas A_i , ce qui permet de propager les segmentations manuelles de A_i sur I . Ayant réalisé cette procédure pour les n atlas, on obtient pour I n labels (c'est à dire n segmentations des organes d'intérêt propagées des n atlas vers I).
- iii) Finalement, une règle de décision est appliquée sur ces n labels afin d'obtenir une seule segmentation pour chaque organe de I . On obtient ainsi une segmentation des organes de I par recalage non-rigide.

Tous les détails concernant cette approche sont disponibles dans [Acosta et al., 2011].

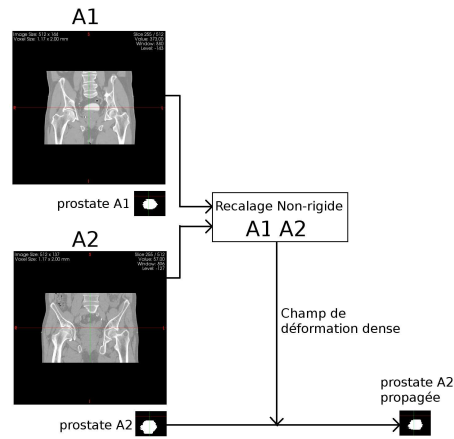
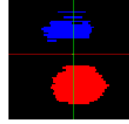


FIGURE 4.33 – Validation croisée « leave-one-out ». Exemple pour 2 atlas A_1 et A_2 . On procède au recalage non-rigide de A_2 (image flottante) sur A_1 (image de référence), ce qui permet d'obtenir un champ dense de déformation non-rigide. Ce champ est ensuite utilisé pour déformer les labels de A_2 (ici la prostate) et les propager dans l'espace de A_1 .

Pour valider cette procédure globale, une procédure de validation croisée de type « leave-one-out » est actuellement menée au laboratoire. Cela doit permettre de mesurer l'impact des méthodes de sélection des atlas utilisées dans iii) ci-dessus. C'est dans ce cadre que se situent les résultats que nous proposons. Dans cette étude, à tour de rôle, chaque atlas A_i tient le rôle du patient I . Après recalage non-rigide entre ce I et les 23 autres atlas, les segmentations manuelles de I sont comparées aux segmentations obtenues par propagation des labels des atlas sur I . Une illustration de cette procédure est donnée sur la figure 4.33, pour 2 atlas A_1 et A_2 . La comparaison est effectuée grâce au coefficient de Dice, dont l'équation (4.1) a été présentée précédemment (4.2). Une illustration des coefficients de Dice obtenus à partir du recalage de la figure 4.33 est donnée figure 4.34. Pour que cette étude soit menée de manière exhaustive, il est nécessaire de recalcr chaque atlas sur les 23 autres, ce qui entraîne pas moins de $24 \times 23 = 552$ recalages non-rigides. Nous donnons dans ce qui suit les résultats préliminaires obtenus.

Dice prostate A1/A2



Dice prostate A1/A2 propagée

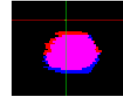


FIGURE 4.34 – illustration du coefficient de Dice. A gauche, le label (prostate) de A_2 (en bleu) ne recouvre pas du tout celui de A_1 (Dice = 0). Après propagation du résultat du recalage non rigide sur le label de A_2 , on obtient un bon recouvrement entre les labels (Dice ≈ 0.8).

Résultats de recalage

Pour étalonner notre procédure de recalage, nous nous sommes concentrés sur seulement quelques-uns des 552 recalages à mener, en ne considérant que les 9 premiers atlas, ce qui représente 72 recalages et nous semblait assez représentatif. Tout comme dans l'étude déjà menée au laboratoire [Dréan et al., 2011], nous nous sommes intéressés seulement aux coefficients de Dice donnés par la prostate, le rectum et la vessie. Pour mener à bien ce travail, le recalage est mené en deux temps :

- un recalage rigide permet d'aligner grossièrement les atlas entre-eux,
- un recalage non-rigide permet de tenir compte des déformations locales.

Pour le recalage rigide, nous avons suivi la même procédure que dans la section 4.7. Nous avons utilisé l'algorithme de descente de gradient stochastique à pas adaptatif avec estimation automatique des paramètres. Nous avons également testé une approche monorésolution et une approche multirésolution avec 5 niveaux de pyramide. De plus, nous avons utilisé 5000 échantillons pour l'estimation de notre mesure de similarité. Les résultats moyens des coefficients de Dice sont donnés dans le tableau 4.11*. Les résultats donnés par l'approche à 5 niveaux de pyramides sont légèrement meilleurs, et nous validons donc cette procédure comme pré-traitement au recalage non-rigide.

Pour le recalage non-rigide, nous avons choisi un modèle de déformation de formes libres (FFD) basé sur les B-Splines présenté au chapitre 1. Nous nous sommes appuyés sur une approche bi-pyramidale pour augmenter progressivement la complexité de données et du modèle de déformation. De plus, nous nous sommes servis de masques binaires pour concentrer nos efforts de recalage sur les zones importantes des images. En effet, si l'on revient par exemple à la figure 4.31 qui présente les données, on constate qu'une grande partie de l'image n'est pas informative, puisque composée de zones noires (l'arrière plan). L'utilisation d'un masque permet de ne pas considérer ces zones pour s'intéresser seulement à l'anatomie du patient. En effet, le masque permet de définir la région dans laquelle seront tirés aléatoirement les pixels rentrant en compte dans le calcul de la mesure de similarité. De plus, nos résultats préliminaires nous ont montré que l'utilisation de masques améliore sensiblement les résultats. Dans cette procédure, les paramètres libres à prendre en compte sont donc :

- le type de masque binaire utilisé,
- le nombre d'échantillons pour l'évaluation de 3EIM. Ce paramètre devient crucial pour le recalage non-rigide, car il influence énormément les temps de calcul.
- le nombre de niveaux à considérer dans la pyramide d'images et dans la pyramide de

transformation, ainsi que les paramètres de ces pyramides,

La première de nos études consiste donc à évaluer le type de masque à utiliser. En effet, pour chaque atlas de la base, nous disposons du masque binaire correspondant. Nous nous sommes donc demandés si il était plus intéressant d'utiliser le masque de l'image de référence, celui de l'image flottante ou une combinaison de deux. Rapidement, l'idée d'utiliser le masque de l'image flottante a été écartée car nous cherchons à transformer cette image flottante pour la mettre dans l'espace de l'image de référence, et les résultats avec un tel masque sont médiocres car on ne peut justement pas se projeter vers l'image de référence. Les masques que nous avons retenus pour l'étude, qui sont donnés sur la figure 4.35, sont le masque de l'image de référence (M_{ref}), ce même masque dilaté par une dilatation morphologique de rayon 5 (MD_{ref}), le masque composé de l'union entre les masques de l'image de référence et de l'image flottante (M_{union}), et ce dernier masque également dilaté (MD_{union}). De plus, nous utilisons ici 1000 échantillons et une approche bi-pyramidale classique, dont nous ne parlerons que plus tard.

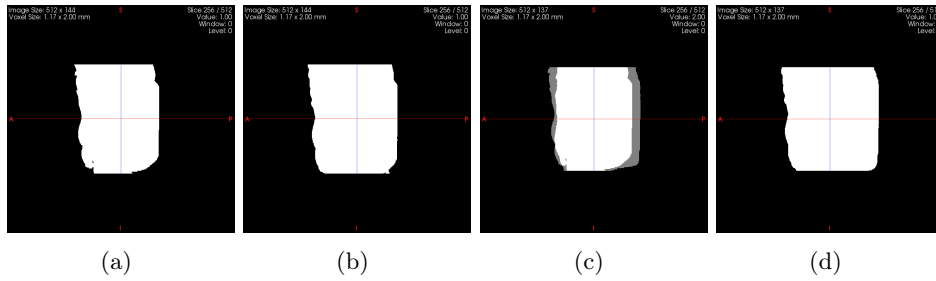


FIGURE 4.35 – Masques utilisés pour le recalage non-rigide. 4.35(a) : masque de l'image de référence. 4.35(b) : masque de l'image de référence dilaté. 4.35(c) : masque d'union de l'image de référence et de l'image flottante. 4.35(d) : masque d'union de l'image de référence et de l'image flottante dilaté.

Les coefficients de Dice moyens obtenus sont donnés dans le tableau 4.12*, et nous ont amenés à utiliser dans la suite le masque de l'image de référence. Ce tableau nous montre également l'amélioration globale des résultats (cf. 4.11*) lorsque l'on considère un recalage non-rigide en plus du recalage rigide initial.

Nous nous sommes ensuite intéressés à l'évolution des coefficients de Dice en fonction du nombre d'échantillons utilisés dans l'estimation de la mesure de similarité. Et nous avons eu la surprise de constater qu'il n'y avait pas forcément de corrélation entre le coefficient et le nombre d'échantillons. Nous donnons par exemple, dans le tableau 4.13, le résultat du recalage des 2 premiers atlas, pour la prostate.

Pourtant visuellement, le recalage semble s'améliorer lorsque l'on augmente le nombre d'échantillons pour l'estimation de 3EIM, comme le montre la figure 4.36*. L'explication est que les zones qui interviennent dans le calcul des coefficients de Dice sont très peu contrastées, et donc peu informatives. Le recalage est donc plutôt conduit par les structures environnantes, comme les os ou le corps du patient. Le recalage des organes d'intérêt se fait en quelque sorte par « défaut ». La déformation de la prostate, par exemple, est plus guidée par les déformations des os l'entourant que par l'information qu'elle porte. Cela nous amène à deux conclusions sur cette étude.

Tout d'abord, notre méthode n'est pas totalement appropriée au problème. Il serait intéressant d'incorporer une information supplémentaire sur les organes d'intérêt afin que ceux-ci soient plus « actifs » dans la procédure de recalage. Cela ouvre de nouvelles perspectives, car l'incorporation de dimensions supplémentaires dans la mesure de similarité est un sujet

de recherche actif (comme nous l'avons vu au chapitre 2). Nous pensons pouvoir, à partir de notre démarche mathématique, développer une nouvelle approximation de l'IM qui prendrait en compte ces nouvelles informations. Une autre approche consisterait à contraindre le recalage sur la zone d'intérêt en utilisant un masque binaire beaucoup plus restrictif, qui cacherait par exemple les vertèbres, fémurs et le contour du corps du patient, qui à l'heure actuelle guident en grande partie le recalage. Cela nécessite un pré-traitement important sur les images pour localiser cette zone d'intérêt et nous ramène en quelque sorte au problème de la segmentation. Mais la piste est néanmoins à envisager.

Le deuxième point, c'est que nous ne pouvons pas réellement nous fier au coefficient de Dice pour évaluer la qualité de notre mesure. En effet, dans toutes les sections précédentes, nous avons vu qu'augmenter le nombre d'échantillons permettait d'obtenir de meilleurs résultats, ce qui n'est plus le cas ici. Pourtant visuellement, tout laisse à penser que le recalage est « meilleur » lorsque plus d'échantillons sont utilisés, mais cela dépend justement de quel sens nous donnons au mot meilleur.

Régler les paramètres de l'approche bi-pyramidale, qui sont les derniers paramètres ajustables, est en conséquence également très difficile. Sans le coefficient de Dice pour juger de la qualité du recalage, nous avons dû nous appuyer sur une inspection visuelle. Ces paramètres sont néanmoins très importants, puisque comme le nombre d'échantillons pris en compte dans le calcul de la mesure de similarité, ils vont jouer énormément sur les temps de calcul nécessaires au recalage. Nous avons donc choisi de régler ces paramètres, à résultat visuel équivalent, en fonction du coût calculatoire.

Pour la pyramide d'images, nous avons choisi pour ces tests d'utiliser une pyramide à 3 niveaux sans sous-échantillonnage, car une pyramide à 4 ou 5 niveaux, pour le recalage non-rigide, devient très difficile à mettre en oeuvre dans des temps de calcul raisonnables. La pyramide de transformation, quant-à-elle, consiste à affiner progressivement la grille de déformation des B-Splines. Pour régler l'écartement entre les Points De Contrôle (PDC) des B-Splines, il est intéressant de se baser sur les dimensions des images à recaler. Les volumes ont une résolution de $512 \times 512 \times 140$ voxels et la taille des voxels est de 1.17×1.17 mm dans le plan de coupe et de 2 mm entre les coupes. Dans nos tests sur les masques, nous avons choisi une pyramide de transformation à 3 niveaux, avec un écartement final entre les PDC de 40 mm, et une multiplication par deux de cet écartement entre les 3 niveaux de la pyramide. C'est à dire qu'à la résolution la plus basse de la pyramide de transformation, les PDC sont écartés de 160 mm. Dans cette configuration, le nombre de paramètres à estimer (nous avons 3 paramètres par PDC, qui correspondent aux composantes du vecteur de translation du PDC) est respectivement de 735, 2541 et 11286 pour chaque niveau de la pyramide. Dans nos tests, nous avons fait varier cet écartement entre les PDC, en regardant l'apport sur la qualité du recalage, et le coût calculatoire. Les résultats que nous donnons ici correspondent à ceux obtenus sur un ordinateur avec un processeur quadri-cœurs cadencé à 2.6 GHz, avec 4Go de RAM. Le tableau 4.14* donne le temps de calcul nécessaire pour toute la procédure de recalage en fonction de l'écartement des PDC et du nombre de paramètres à estimer à la plus haute résolution, ces 2 paramètres étant complètement liés. Ces résultats représentent une moyenne sur 20 recalages et ont été obtenus en considérant 1000 échantillons dans l'estimation de la mesure de similarité. De plus, nous ne donnons ici que les 3 jeux de paramètres qui nous ont donné des recalages visuellement satisfaisants. Lorsque l'on réduit encore l'écartement entre les PDC, des déformations aberrantes ont tendance à se produire, comme illustré sur la figure 4.37*. Quand au contraire l'écartement augmente, l'image flottante n'est pas assez déformée pour correspondre à l'image de référence.

Parmi ces 3 jeux de paramètres, nous avons finalement conservé un écartement final de 40 mm entre les PDC, puisque c'est cette configuration qui entraîne la charge calculatoire la plus

faible. Cependant, nous avons initialement choisi cette base de données car elle donnait accès au coefficient de Dice permettant en théorie de quantifier la qualité du recalage. En pratique, nous avons fini par nous attacher aux résultats visuels qui donnent seulement une idée qualitative approximative. Partant de ce constat, nous avons décidé de tester notre algorithme sur une autre base de données, disponible sur internet, et qui donne 4 scores différents pour chaque recalage. Cette base est présentée dans ce qui suit. Concernant cette partie, nous avons continué les tests, car les résultats semblaient quand même relativement bons, sur les premiers atlas de la base, par rapport à ceux obtenus au laboratoire [Dréan et al., 2011]. Nous avons donc mené à terme l'étude, en effectuant les 552 recalages nécessaires à la validation croisée. A partir de tous les tests préliminaires que nous venons d'évoquer, nous avons choisi comme prétraitement un recalage rigide prenant en compte 5000 échantillons dans l'estimation de la mesure de similarité, couplé à une pyramide d'images à 5 niveaux avec au maximum 1000 itérations de l'algorithme à chaque niveau de la pyramide. Ensuite, nous avons choisi le recalage non-rigide basé sur les B-Splines en utilisant 500 échantillons dans l'évaluation de la mesure de similarité. De plus, nous nous sommes aidés du masque binaire de l'image de référence qui semblait donner les meilleurs résultats. Enfin, nous avons utilisé une approche bi-pyramidale avec 3 niveaux dans la pyramide d'images et dans la pyramide de transformation, et un écartement final entre les PDC de 40 mm. Les résultats obtenus sont représentés synthétiquement sur la figure 4.38*. Comme nous l'avons déjà évoqué précédemment, notre approche n'est pas totalement appropriée au problème, c'est pourquoi nous avons considéré une seconde base de données que nous présentons à présent.

4.8.2 Base de données EMPIRE

La deuxième base de données permettant de valider notre mesure de similarité pour une utilisation en recalage non-rigide est issue du challenge MICCAI 2010 nommé EMPIRE (Evaluation of Methods for Pulmonary Image Registration). Elle est constituée de 30 paires de volumes scanner du thorax. Chaque paire est prise d'un même sujet, donc il s'agit ici de recalage intra-sujet. Les images proviennent de plusieurs imageurs et instituts de recherche différents, et n'ont donc pas toutes la même résolution spatiale ni la même taille de voxels. Les scanners ont pu être enregistrés à différents moments du cycle respiratoire. Les patients peuvent avoir une pathologie pulmonaire ou être sains. En plus de ces données, des masques binaires des poumons sont disponibles. Ils ont été générés automatiquement en utilisant un algorithme proposé par [van Dongen and van Ginneken, 2010]. Ils ont été vérifiés visuellement et corrigés si nécessaire. Deux exemples d'images et de leurs masques correspondants sont donnés sur la figure 4.39*. On voit bien la grande variabilité des images suivant les imageurs et les patients.

Cette base nous a intéressé car elle permet l'évaluation de la qualité du recalage non-rigide par la soumission des résultats de recalage en ligne. De plus, les résultats étant disponibles sur le site, cela permet une comparaison de notre méthode avec celles proposées par les autres participants. La qualité du recalage est évaluée de 4 manières différentes, comme nous l'avons précisé au début de ce chapitre 4.2.

La soumission des résultats consiste à envoyer les champs de déformation denses issus des recalages, et ce pour les 30 patients. Cela donne une difficulté supplémentaire pour le test de cette base de données. En effet, pour fixer les paramètres de notre méthode de recalage, nous ne pouvons pas cette fois nous appuyer sur un score, comme c'était le cas dans la partie précédente, puisque cela nécessiterait la soumission de la méthode en ligne avant d'avoir réglé ses paramètres. Nous avons donc simplement utilisé des résultats visuels pour ce faire. De plus, la question qui peut se poser est de savoir quel niveau de précision est attendu dans cette

étude, car encore une fois il faut toujours considérer le compromis entre temps de calcul et précision du résultat. Nous avons donc réalisé des tests préliminaires qui nous ont permis de sélectionner les paramètres optimaux pour effectuer les recalages sur cette base de données. Le prétraitement retenu ici est un recalage affine, car les tests préliminaires ont montré de meilleurs résultats que ceux obtenus après un recalage rigide. Les paramètres auxquels nous nous sommes attachés ici sont :

- les paramètres des pyramides d’images et de transformation. Nous avons sélectionné 4 niveaux dans les 2 pyramides, avec un écartement final entre les PDC de 20 mm.
- le nombre d’échantillons pour l’évaluation de 3EIM. Nous en avons utilisé 2000 lors de la procédure de recalage affine, et 1000 pour le recalage non-rigide.
- le nombre d’itérations à chaque niveau de la pyramide. Il a été fixé à 1000 pour le recalage affine, et à 200 pour la procédure non-rigide.

Quelques exemples visuels des résultats que nous avons obtenus lors de la phase de paramétrisation sont donnés sur les figures 4.40* et 4.41*. Celles-ci montrent la difficulté à se baser sur ces seuls résultats visuels, les différences concernant l’évaluation de la qualité du recalage étant parfois difficilement discernables à l’œil nu. Néanmoins, nous avons mené les recalages sur les 30 volumes de la base, en utilisant les paramètres précédemment cités, et avons soumis les résultats sous formes de champs de déformation denses. La totalité des résultats obtenus est disponible à l’adresse <http://empire10.isi.uu.nl/mainResults.php> (sous le nom d’équipe « 3-EMI LTSI »). Avec des tels résultats, nous sommes classés 24ème, sur 27 équipes ayant participé au challenge. Les résultats sont encourageants comparés aux autres équipes, car beaucoup d’entre-elles proposent une méthode de recalage spécifiquement dédiée au problème posé, alors que notre méthode se veut avant tout générique. Parmi ces équipes, beaucoup utilisent le même type de procédure de recalage (c’est à dire un modèle de déformation de type B-Spline, avec l’utilisation de l’IM comme mesure de similarité), en venant y adjoindre une énergie ou un prétraitement qui permet de contraindre le champ de déformation et ainsi d’obtenir de meilleurs résultats. Nous pensons qu’en ajoutant à notre méthode une telle énergie, il serait également possible d’obtenir de meilleurs résultats.

Nous présentons dans les tableaux 4.42*, 4.43*, 4.44* et 4.45* les résultats détaillés par rapport aux quatre critères d’évaluation dont nous avons parlé dans la section 4.2. Ceux-ci méritent d’être commentés plus en détail :

- le tableau 4.42* présente les résultats de l’alignement des frontières des poumons. Nous pouvons voir que pour 2 patients (05 et 25), le résultat est parfait au sens du critère. Pour 22 autres patients, les résultats sont très bons (ils se situent dans les mêmes ordres de grandeur que les résultats des meilleures équipes ayant participé au challenge), tandis que pour les 6 derniers patients (06, 07, 14, 20, 28, 30), les résultats sont moins bons. Finalement, ces 6 derniers patients influent grandement sur le résultat global. En utilisant une méthode moins rapide mais plus précise, il aurait sans aucun doute été possible d’obtenir de très bons résultats pour ces derniers patients. Nous voyons que le compromis précision/ temps de calcul est encore une fois primordial pour les applications de recalage non-rigide.
- de la même manière, dans le tableau 4.43* qui présente les résultats d’alignement des scissures principales, nous constatons 7 résultats parfaits, 11 très bons résultats et 12 résultats moyens ou médiocres. Les conclusions à tirer sont les mêmes que précédemment.
- pour le critère de correspondance entre points annotés (tableau 4.44*), nous obtenons 19 bons résultats et 11 résultats moyens.
- enfin, en ce qui concerne les singularités dans le champ de déformation, nous n’obtenons que des résultats parfaits au sens du critère (tableau 4.45*). Le choix des B-Splines comme modèle de déformation non-rigide est en partie responsable de ce bon résultat,

puisque c'est un modèle de déformation intrinsèquement régulier. Cependant, si il est mal paramétré, il peut conduire à des résultats erronés et à des champs de déformation aberrants.

A travers cette évaluation, nous avons donc montré le potentiel de notre nouvelle mesure de similarité 3EIM pour le recalage non-rigide monomodal. A présent, d'autres pistes sont envisageables pour améliorer les résultats de cette mesure, comme l'utilisation d'a priori sur les images à recaler, qui permettrait de contraindre le champ de déformation et d'obtenir ainsi des déformations encore plus réalistes.

4.9 Conclusion

Au cours de ce chapitre, nous avons intensivement évalué nos différentes mesures de similarité, et les avons comparé à différentes estimées de l'IM de référence dans le domaine du recalage d'images médicales multimodales.

Pour commencer, nous avons considéré ces mesures en dehors de tout contexte de recalage, pour mettre en avant leurs qualités et leurs défauts propres. Cela nous a permis d'écarter l'IM estimée par histogramme. A ce stade, nous avons également fait le choix de ne conserver qu'une des 6 mesures de similarité que nous proposons, en l'occurrence 3EIM. Ce choix a été fait pour gagner du temps dans le paramétrage de notre mesure. En effet, nous pensons par exemple que 3EIMN doit avoir un comportement très similaire, et que les tests que nous avons menés au cours de cette thèse seront facilement transposables à cette seconde mesure. Pour ce qui est de nos mesures basées sur un développement d'Edgeworth à l'ordre 4, un travail important est encore nécessaire pour comprendre leur comportement atypique, mais nous pensons que sa régularité démontrée peut permettre d'en faire une mesure très efficace pour le recalage. Des tests complémentaires sont néanmoins nécessaires.

Ceci étant fait, nous avons inséré notre mesure de similarité dans un schéma global de recalage, en testant notamment les optimiseurs de référence donnés dans la littérature. Pour ce faire, nous avons utilisé les données d'entraînement de la base RIRE, qui permet d'obtenir une évaluation quantitative précise de la qualité du recalage. Une fois l'optimiseur choisi et paramétré, nous avons utilisé les données de test de cette même base pour évaluer l'utilisation de notre mesure de similarité pour des tâches de recalage rigide multimodal. Nous avons constaté que notre mesure permettait un recalage rigide efficace, notamment quand on considère des images multimodales TEP/IRM.

Enfin, nous avons voulu évaluer notre mesure sur des tâches de recalage non-rigide, puisque c'est la finalité de ce travail de thèse. Les résultats préliminaires ont été très encourageants, mais le premier jeu de données testé (la base de données CT du pelvis du laboratoire) ne permet pas une analyse fine de la qualité du recalage. Aussi, nous nous sommes intéressés à une seconde base de données (EMPIRE) qui a donné des résultats quantitativement plus parlants. Bien que notre méthode n'était pas spécifiquement adaptée à ce nouveau jeu de données, nous avons obtenu des résultats très satisfaisant pour plus de la moitié des patients de la base. Ici de nouvelles pistes sont envisageables pour améliorer la qualité du recalage sur les patients dont les résultats sont moins bons. Il serait ainsi possible d'apporter une information supplémentaire à celle d'intensité afin de rendre le recalage plus spécifique aux données. L'ajout d'une information spatiale est par exemple envisageable. Enfin, nous avons constaté tout au long de cette étude que le paramétrage de notre méthode pour le recalage non-rigide requiert des compromis entre la précision et les temps de calcul, qui peuvent devenir rapidement très importants.

Bibliographie

- [Acosta et al., 2011] Acosta, O., Simon, A., Monge, F., Commandeur, F., Bassirou, C., Cazoulat, G., de Crevoisier, R., and Haigron, P. (2011). Evaluation of multi-atlas-based segmentation of ct scans in prostate cancer radiotherapy. *IEEE International Symposium on Biomedical Imaging (ISBI)*.
- [Bardera et al., 2006] Bardera, A., Feixas, M., Boada, I., and Sbert, M. (2006). High dimensional normalized mutual information for image registration using random lines. In *International Workshop on Medical Image Registration*.
- [Cahill et al., 2009] Cahill, N., Schnabel, J., Noble, J., and Hawkes, D. (2009). Overlap invariance of cumulative residual entropy measures for multimodal image alignment. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series*, volume 7259.
- [Collignon et al., 1995] Collignon, A., Maes, F., Delaere, D., Vandermeulen, D., Suetens, P., and Marchal, G. (1995). Automated Multi-modality Image Registration based on Information Theory. In *Conference on Information Processing in Medical Imaging*, pages 263–274.
- [Dai, 2003] Dai, Y.-H. (2003). A family of hybrid conjugate gradient methods for unconstrained optimization. *Mathematics of Computation*, 72 :1317–1328.
- [Dréan et al., 2011] Dréan, G., Monge, F., Simon, A., Cazoulat, G., De Crevoisier, R., Haigron, P., and Acosta, O. (2011). Evaluation de méthodes de recalage dense basées sur l'intensité de ct-scans dans le cadre du traitement du cancer de la prostate par radiothérapie guidée par l'image. *Recherche en Imagerie et Technologie pour la Santé (RITS)*.
- [Škerl et al., 2006] Škerl, D., Likar, B., and Pernuš, F. (2006). Evaluation of similarity measures for non-rigid registration. *Biomedical Image Registration, Lecture Notes in Computer Science*, 4057 :160–168.
- [Kabus et al., 2009] Kabus, S., Klinder, T., Murphy, K., Ginneken, B., Lorenz, C., and Pluim, J. P. (2009). Evaluation of 4D-CT lung registration. *Proceedings of the 12th International Conference on Medical Image Computing and Computer-Assisted Intervention : Part I*, pages 747–754.
- [Klein et al., 2009] Klein, S., Pluim, J., Staring, M., and Viergever, M. (2009). Adaptive stochastic gradient descent optimisation for image registration. *International Journal of Computer Vision*, 81(3) :227–239.
- [Klein et al., 2010] Klein, S., Staring, M., Murphy, K., Viergever, M., and Pluim, J. (2010). elastix : a toolbox for intensity-based medical image registration. *IEEE Transactions on Medical Imaging*, 29(1) :196 – 205.
- [Klein et al., 2007] Klein, S., Staring, M., and Pluim, J. P. W. (2007). Evaluation of optimization methods for nonrigid medical image registration using mutual information and b-splines. *IEEE Transactions on Image Processing*, 16(12) :2879–2890.
- [Lester and Arridge, 1999] Lester, H. and Arridge, S. (1999). A survey of hierarchical non-linear medical image registration. *Pattern Recognition*, 32(1) :129–149.
- [Mattes et al., 2003] Mattes, D., Haynor, D. R., Vesselle, H., Lewellen, T. K., and Eubank, W. (2003). PET-CT image registration in the chest using free-form deformations. *IEEE Transactions on Medical Imaging*, 22(1) :120–128.
- [Murphy et al., 2008] Murphy, K., van Ginneken, B., Pluim, J., Klein, S., and Staring, M. (2008). Semi-automatic reference standard construction for quantitative evaluation of lung ct registration. In *Medical Image Computing and Computer Assisted Intervention MICCAI 2008*, volume 5242 of *Lecture Notes in Computer Science*. Springer Berlin / Heidelberg.

- [Pluim et al., 2001] Pluim, J., Maintz, J., and Viergever, M. (2001). f-Information measures in medical image registration. In Sonka, M. and Hanson, K., editors, *SPIE Medical Imaging*, volume 4322, pages 579–587. SPIE press.
- [Studholme et al., 1997] Studholme, C., Hill, D. L. G., and Hawkes, D. J. (1997). Automated three-dimensional registration of magnetic resonance and positron emission tomography brain images by multiresolution optimization of voxel similarity measures. *Medical Physics*, 24(1) :25–35.
- [Studholme et al., 1999] Studholme, C., Hill, D. L. G., and Hawkes, D. J. (1999). An overlap invariant entropy measure of 3D medical image alignment. *Pattern Recognition*, 32(1) :71–86.
- [Thévenaz and Unser, 1997] Thévenaz, P. and Unser, M. (1997). Spline pyramids for inter-modal image registration using mutual information. In *Proceedings of the SPIE Conference on Mathematical Imaging : Wavelet Applications in Signal and Image Processing V*, volume 3169, pages 236–247, San Diego CA, USA.
- [van Dongen and van Ginneken, 2010] van Dongen, E. and van Ginneken, B. (2010). Automatic segmentation of pulmonary vasculature in thoracic CT scans with local thresholding and airway wall removal. *Proceedings of the 2010 IEEE international conference on Biomedical imaging : from nano to Macro*, pages 668–671.
- [van Rikxoort et al., 2009] van Rikxoort, E., de Hoop, B., Viergever, M., Prokop, M., and van Ginneken, B. (2009). Automatic lung segmentation from thoracic computed tomography scans using a hybrid approach with error detection. *Medical Physics*, 36(7) :2934–2947.
- [van Rikxoort et al., 2008] van Rikxoort, E., van Ginneken, B., Klik, M., and Prokop, M. (2008). Supervised enhancement filters : application to fissure detection in chest CT scans. *IEEE Transactions on Medical Imaging*, 27(1) :1–10.
- [Viola, 1995] Viola, P. (1995). *Alignment by Maximization of Mutual Information*. PhD thesis, Massachusetts Institute of Technology Artificial Intelligence Laboratory, Boston, USA.
- [Viola and Wells, 1997] Viola, P. and Wells, W. (1997). Alignment by maximization of Mutual Information. *International Journal of Computer Vision*, 24(2) :137–154.
- [West et al., 1997] West, J., Fitzpatrick, J. M., Wang, M. Y., Dawant, B. M., Maurer, C. R., Kessler, R. M., Maciunas, R. J., Barillot, C., Lemoine, D., Collignon, A., Maes, F., Sumanaweera, T. S., Harkness, B., Hemler, P. F., Hill, D. L. G., Hawkes, D. J., Studholme, C., Maintz, J. B. A., Viergever, M. A., Mal, G., Pennec, X., Noz, M. E., Maguire, G. Q., Pollock, M., Pelizzari, C. A., Robb, R. A., Hanson, D., and Woods, R. P. (1997). Comparison and Evaluation of Retrospective Intermodality Image Registration Techniques. *Journal of Computer Assisted Tomography*, 21(4) :554–566.

TABLE 4.3 – Paramètres des différents optimiseurs testés.

Algorithme		G	GC	GS_auto	GS_manu	LBFGS
Choix du pas	PI	4, 8, 16	0.5, 1, 2			0.5, 1, 2
	FR	0.1, 0.5, 0.9				
	c_1		0.001, 0.0001			0.001, 0.0001
	c_2		0.7, 0.9			0.7, 0.9
	ItLS		5, 10, 20			5, 10, 20
	a				200, 400, 1000, 2000	
	A			10, 20, 50	10, 20, 50	
	α				0.5, 1, 1.5	
	f_{\min}				-0.8, 0.5	
	f_{\max}				0.5, 1	
	ω				0.00001, 0.0000001	
	t_0			0.0, 0.5	0.0, 0.5	
Direction de recherche	c_k	0.01, 0.1, 1				
	ζ_k		FR, PR, HS, DY, DYHS*			
	M					5, 10, 20
Critères d'arrêt	ItMax	200, 500, 1000				
	Pmin	0.001, 0.0001, 0.00001				
	Gmin	0.001, 0.0001, 0.00001				
	Wolfe		Vrai, Faux			Vrai, Faux
Schéma multirésolution		avec ou sans sous-échantillonnage, 1, 2, ou 3 niveaux de pyramide				
*Fletcher-Reeves, Polak-Ribiere, Hestenes-Stiefel, Dai-Yuan, Dai-Yuan-Hestenes-Stiefel, d'après [Dai, 2003, Klein et al., 2007]						

TABLE 4.4 – Pourcentage des recalages aboutis pour les différents optimiseurs, avec la mesure E3IM.

Modalité 1	Modalité 2	G	GC	GS_manu	GS_auto	LBFGS
CT	IRM PD	88.00%	90.00%	92.00%	100.00%	100.00%
CT	IRM PD rectifié	85.00%	95.00%	97.50%	100.00%	100.00%
CT	IRM T1	93.00%	98.00%	100.00%	100.00%	100.00%
CT	IRM T1 rectifié	96.00%	98.50%	100.00%	100.00%	100.00%
CT	IRM T2	93.00%	99.50%	100.00%	100.00%	100.00%
CT	IRM T2 rectifié	99.00%	99.50%	100.00%	100.00%	100.00%
TEP	IRM PD	93.00%	99.50%	99.50%	100.00%	100.00%
TEP	IRM PD rectifié	87.00%	100.00%	99.50%	100.00%	100.00%
TEP	IRM T1	93.00%	100.00%	100.00%	100.00%	100.00%
TEP	IRM T1 rectifié	99.00%	100.00%	100.00%	100.00%	100.00%
TEP	IRM T2	100.00%	99.00%	99.50%	100.00%	100.00%
TEP	IRM T2 rectifié	97.00%	100.00%	100.00%	100.00%	100.00%

TABLE 4.5 – Pourcentage des recalages dont l'erreur moyenne est inférieure à 5mm, avec la mesure E3IM.

Modalité 1	Modalité 2	G	GC	GS_manu	GS_auto	LBFGS
CT	IRM PD	1.00%	19.00%	31.50%	89.00%	0.00%
CT	IRM PD rectifié	21.00%	5.50%	29.50%	83.00%	24.00%
CT	IRM T1	11.00%	22.00%	37.00%	80.00%	0.00%
CT	IRM T1 rectifié	32.00%	4.50%	36.00%	89.00%	25.00%
CT	IRM T2	3.50%	2.50%	42.00%	87.00%	0.00%
CT	IRM T2 rectifié	24.00%	3.00%	39.50%	89.00%	76.00%
TEP	IRM PD	32.00%	96.50%	71.50%	100.00%	50.00%
TEP	IRM PD rectifié	34.00%	95.00%	74.00%	100.00%	100.00%
TEP	IRM T1	27.00%	96.00%	67.00%	96.00%	100.00%
TEP	IRM T1 rectifié	37.50%	77.00%	46.00%	96.00%	100.00%
TEP	IRM T2	54.00%	93.00%	52.00%	100.00%	100.00%
TEP	IRM T2 rectifié	38.30%	4.00%	18.00%	96.00%	0.00%

TABLE 4.6 – Temps de calcul moyen, en secondes, pour 1000 itérations du recalage rigide 3D, en fonction de la mesure de similarité employée.

	Nombre d'échantillons			
Mesures de similarité	50	100	500	1000
3EIM	2.2	2.8	6.1	12.1
IM _{BS}	0.45	0.61	1.9	3.8
IMN _{BS}	0.45	0.53	1.6	2.9
IM _G	3	5	107	419

TABLE 4.7 – Données disponibles selon le patient. Base RIRE.

[illegible]

TABLE 4.8 – Erreur moyenne de recalage (en mm), pour les différents couples de modalité d'imagerie, et pour les différentes mesures de similarité, en fonction du nombre d'itérations maximal retenu (200 ou 1000). Résultat sur les 9 premiers patients de la base de données RIRE.

		3EIM		IM _{BS}		IMN _{BS}		IM _G	
M1	M2	200	1000	200	1000	200	1000	200	1000
CT	IRM_PD	5.232	5.382	11.169	11.167	5.72	5.721	17.403	19.634
CT	IRM_PD_rectifié	6.196	6.33	10.093	10.065	10.963	10.962	19.765	18.782
CT	IRM_T1	5.959	6.305	13.677	13.677	13.719	13.727	18.961	20.395
CT	IRM_T1_rectifié	5.003	5.206	7.621	7.618	6.011	5.458	23.614	19.8
CT	IRM_T2	6.485	5.29	15.935	15.937	15.036	15.252	21.473	22.711
CT	IRM_T2_rectifié	5.563	5.561	11.958	11.959	9.11	9.11	20.225	22.127
TEP	IRM_PD	4.534	4.208	6.514	6.505	6.248	6.255	13.371	10.43
TEP	IRM_PD_rectifié	4.056	4.18	5.58	5.589	4.984	4.967	11.098	9.11
TEP	IRM_T1	5.012	4.986	6.939	6.94	6.747	6.746	11.177	13.157
TEP	IRM_T1_rectifié	4.513	4.324	5.165	5.172	5.034	5.041	5.047	8.625
TEP	IRM_T2	4.225	4.313	9.001	9.002	9.221	9.22	10.621	12.238
TEP	IRM_T2_rectifié	3.953	3.979	5.624	5.624	5.079	5.072	11.84	10.632
		Essai8_1	Essai8_4	Essai_4_2	Essai5_2	Essai_4_1	Essai5_1	Essai_4_4	Essai_4_3

TABLE 4.9 – Apport du nombre de niveaux de la pyramide d’images sur l’erreur moyenne de recalage.

Patient 001		Niveaux de la pyramide				
Modalité 1	Modalité 2	1	2	3	4	5
CT	IRM_PD	3.869	4.782	4.198	3.853	3.617
CT	IRM_PD_rectifié	3.6	3.101	3.23	3.566	3.393
CT	IRM_T1	3.52	4.674	3.248	3.437	3.17
CT	IRM_T1_rectifié	3.445	3.411	3.814	3.792	3.684
CT	IRM_T2	5.759	5.132	4.829	4.972	4.674
CT	IRM_T2_rectifié	6.135	2.228	2.028	3.89	3.78
		test	Essai6_1	Essai6_2	Essai6_3	Essai6_4

TABLE 4.10 – Erreur moyenne de recalage (en mm), pour les différents couples de modalité d’imagerie, et pour les différentes mesures de similarité, avec l’utilisation d’une pyramide d’images à 5 niveaux.

Modalité 1	Modalité 2	3EIM	IM _{BS}	IMN _{BS}	IM _G
CT	IRM_PD	6.236	5.543	5.114	25.68
CT	IRM_PD_rectifié	5.03	4.808	4.867	26.747
CT	IRM_T1	5.459	6.686	5.932	36.282
CT	IRM_T1_rectifié	5.661	7.742	6.243	29.504
CT	IRM_T2	5.664	6.244	5.74	30.2
CT	IRM_T2_rectifié	5.777	5.349	5.002	33.976
TEP	IRM_PD	5.206	9.299	8.158	19.61
TEP	IRM_PD_rectifié	5.122	5.55	7.399	20.473
TEP	IRM_T1	5.378	9.692	7.963	28.218
TEP	IRM_T1_rectifié	4.813	12.372	7.89	24.63
TEP	IRM_T2	6.931	10.64	7.441	25.343
TEP	IRM_T2_rectifié	7.156	8.092	7.824	28.018
		Essai7_1	Essai7_2	Essai7_3	Essai7_4

TABLE 4.11 – Coefficients de Dice moyens obtenus sur les 9 premiers patients de la base, en fonction de la procédure de recalage rigide utilisée.

organe Procédure de recalage	prostate	vessie	rectum
1 niveau	0.24	0.39	0.22
5 niveaux	0.25	0.43	0.23

TABLE 4.12 – Coefficients de Dice moyens obtenus sur les 9 premiers patients de la base, en fonction du type de masque utilisé.

Masque utilisé \ organe	prostate	vessie	rectum
M_{ref}	0.59	0.56	0.49
MD_{ref}	0.56	0.55	0.47
M_{union}	0.53	0.51	0.45
MD_{union}	0.53	0.51	0.47

TABLE 4.13 – Coefficient de Dice de la prostate, pour le recalage des atlas 1 et 2, en fonction du nombre d'échantillons

Nombre d'échantillons	200	500	1000	2000	5000
Dice de la prostate	0.82229	0.691757	0.689241	0.773345	0.596537

TABLE 4.14 – Temps nécessaire pour une procédure de recalage non-rigide en fonction de l'écartement (en mm) entre les points de contrôle et du nombre de paramètres à estimer au niveau le plus haut de la pyramide. Nous considérons ici un écartement égal suivant les 3 dimensions de l'image.

Ecartement (en mm) entre les PDC	Nombre de paramètres	Temps d'exécution (min)
20	60588	384
30	21528	134
40	11286	69

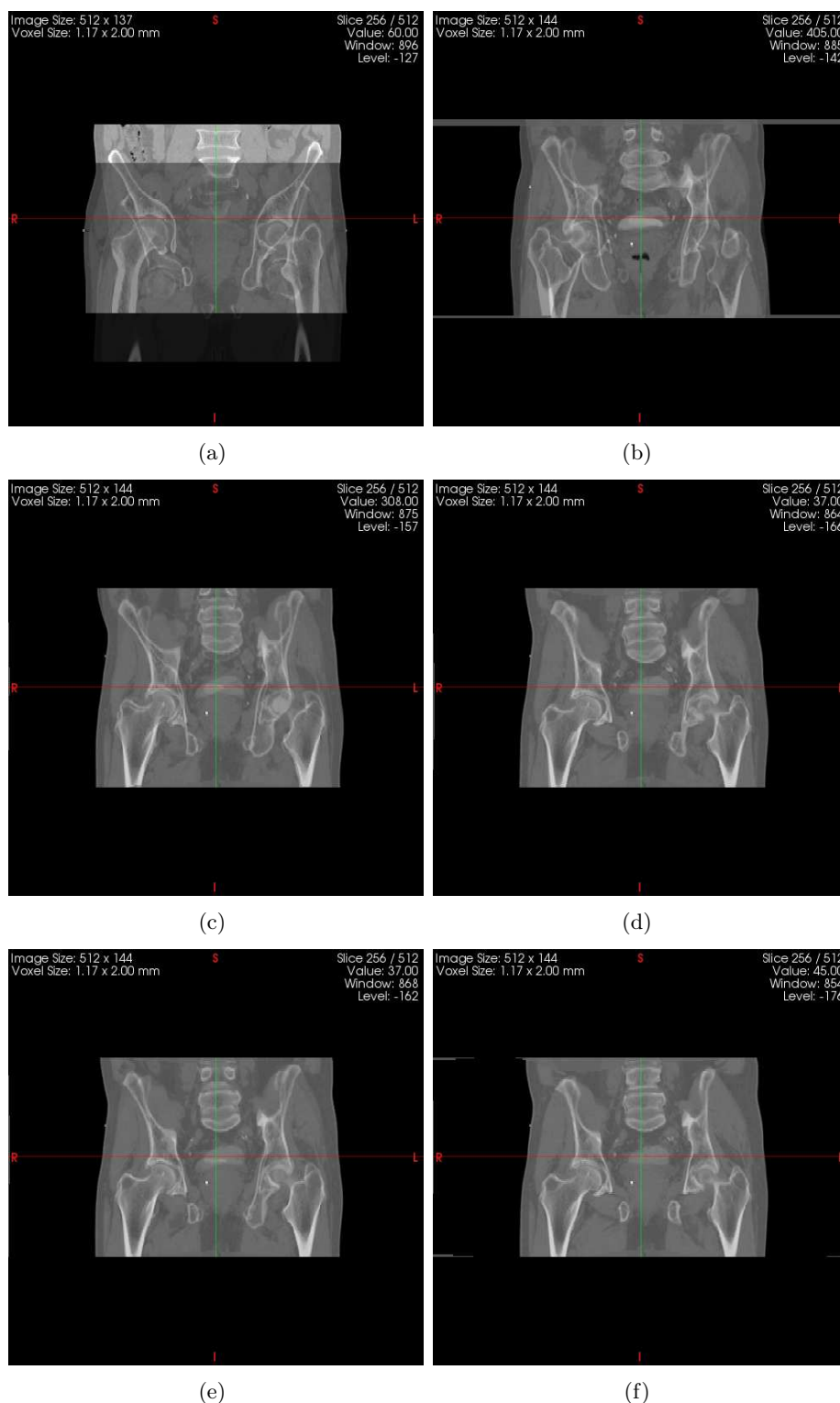


FIGURE 4.36 – Evaluation visuelle par superposition de la qualité du recalage en fonction du nombre d'échantillons utilisés dans l'estimation de 3EIM. 4.36(a) Images avant recalage. 4.36(b) Recalage rigide. 4.36(c) Recalage non-rigide avec 200 échantillons. 4.36(d) Recalage non-rigide avec 500 échantillons. 4.36(e) Recalage non-rigide avec 1000 échantillons. 4.36(f) Recalage non-rigide avec 2000 échantillons.

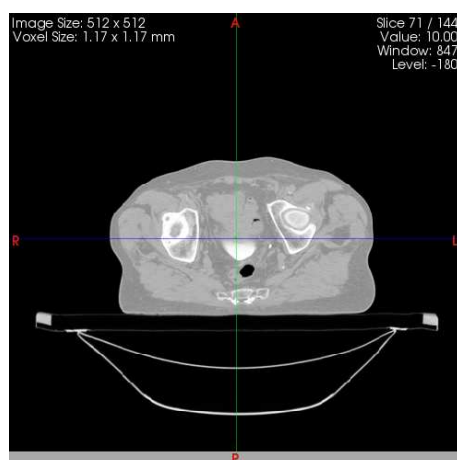


FIGURE 4.37 – Exemple de résultats de recalage non-rigide aberrant. Le contour du patient présente des sortes d’ « ondulations » dues à l’écart trop faible entre les points de contrôle.

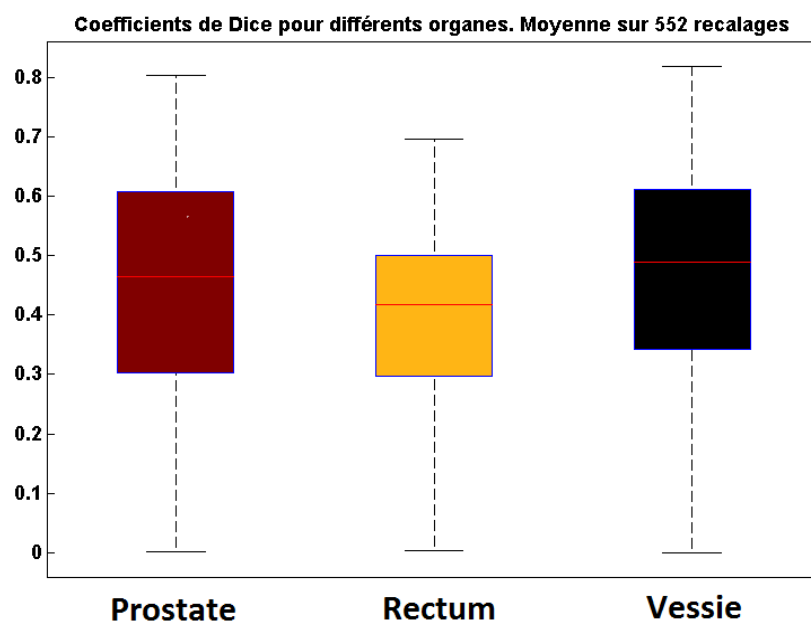


FIGURE 4.38 – Vue d’ensemble des coefficients de Dice obtenus lors de la validation croisée, pour les différents organes considérés. Ces résultats représentent la moyenne sur les 552 recalages effectués.

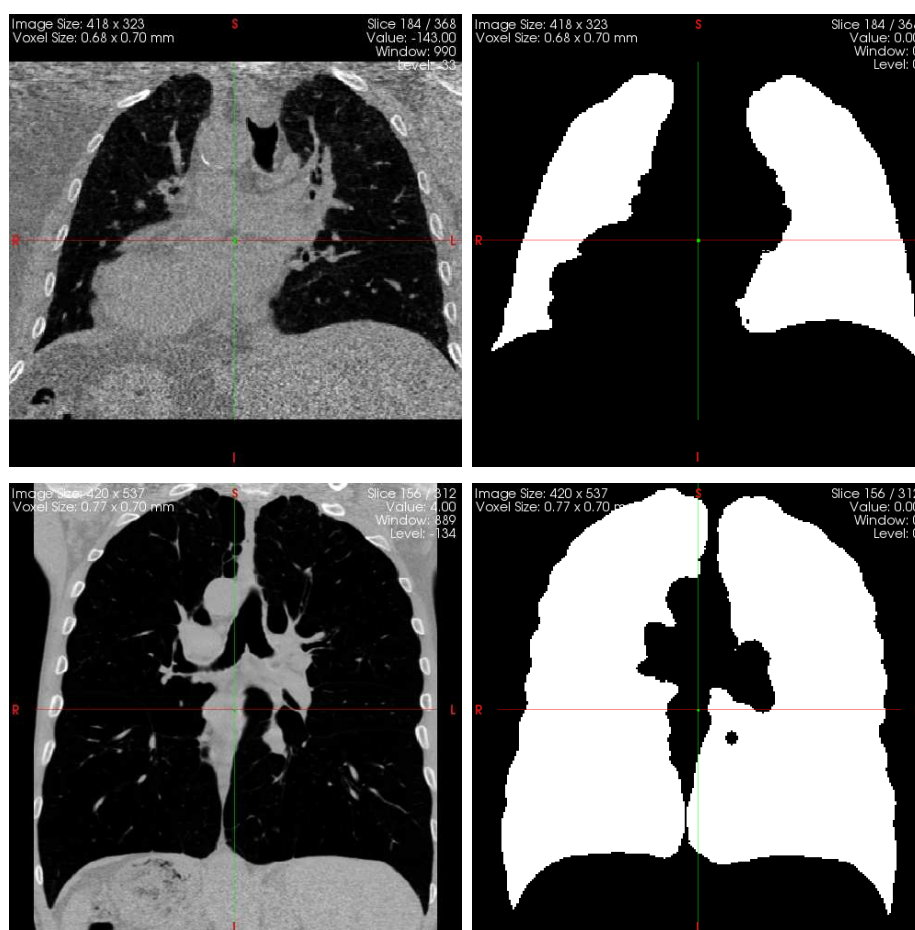


FIGURE 4.39 – Exemples d’images de la base EMPIRE, et des masques binaires des poumons correspondants.

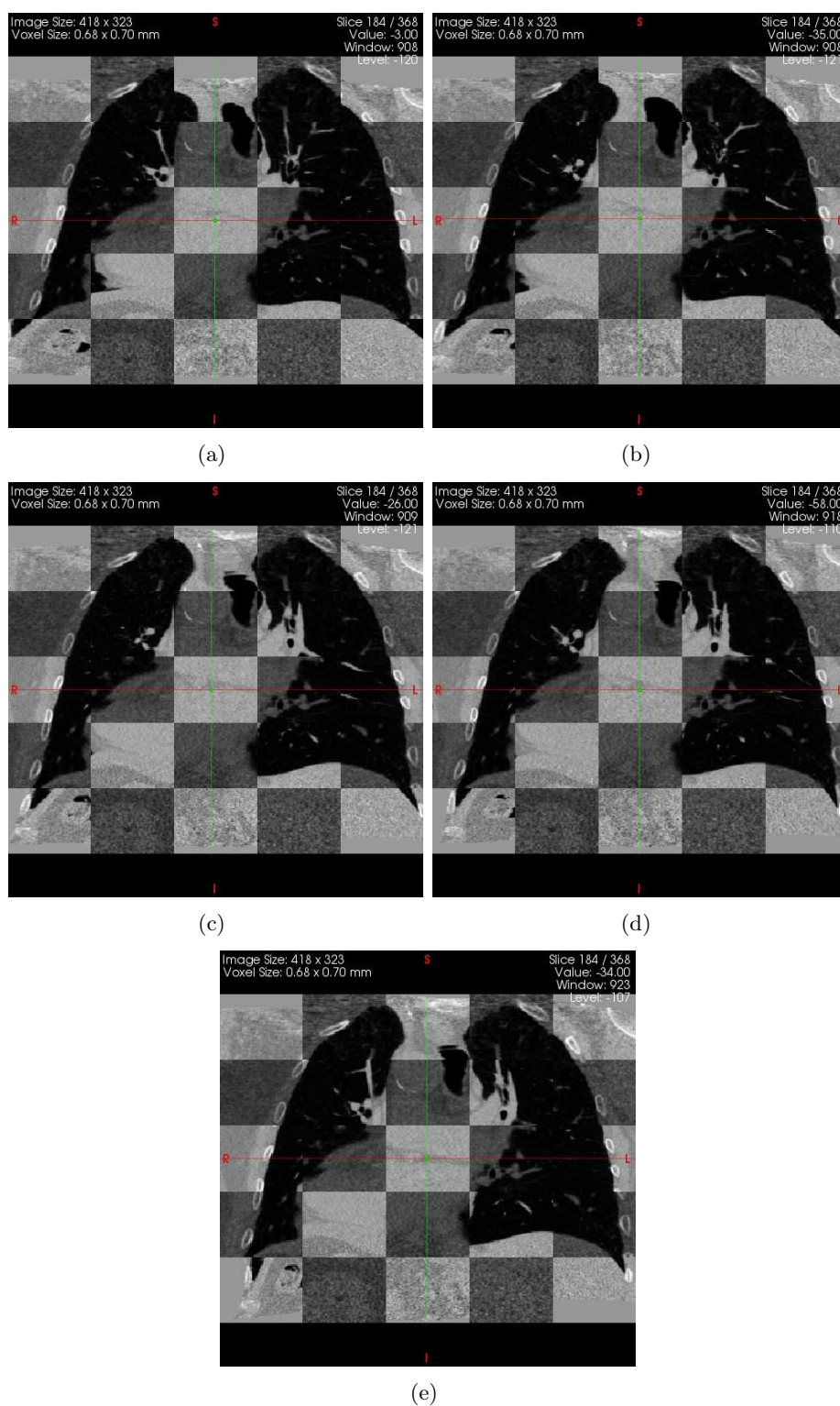


FIGURE 4.40 – Evaluation visuelle de la qualité des recalages en fonction du nombre d'échantillons, pour respectivement 50, 100, 300, 600 et 1000 échantillons.

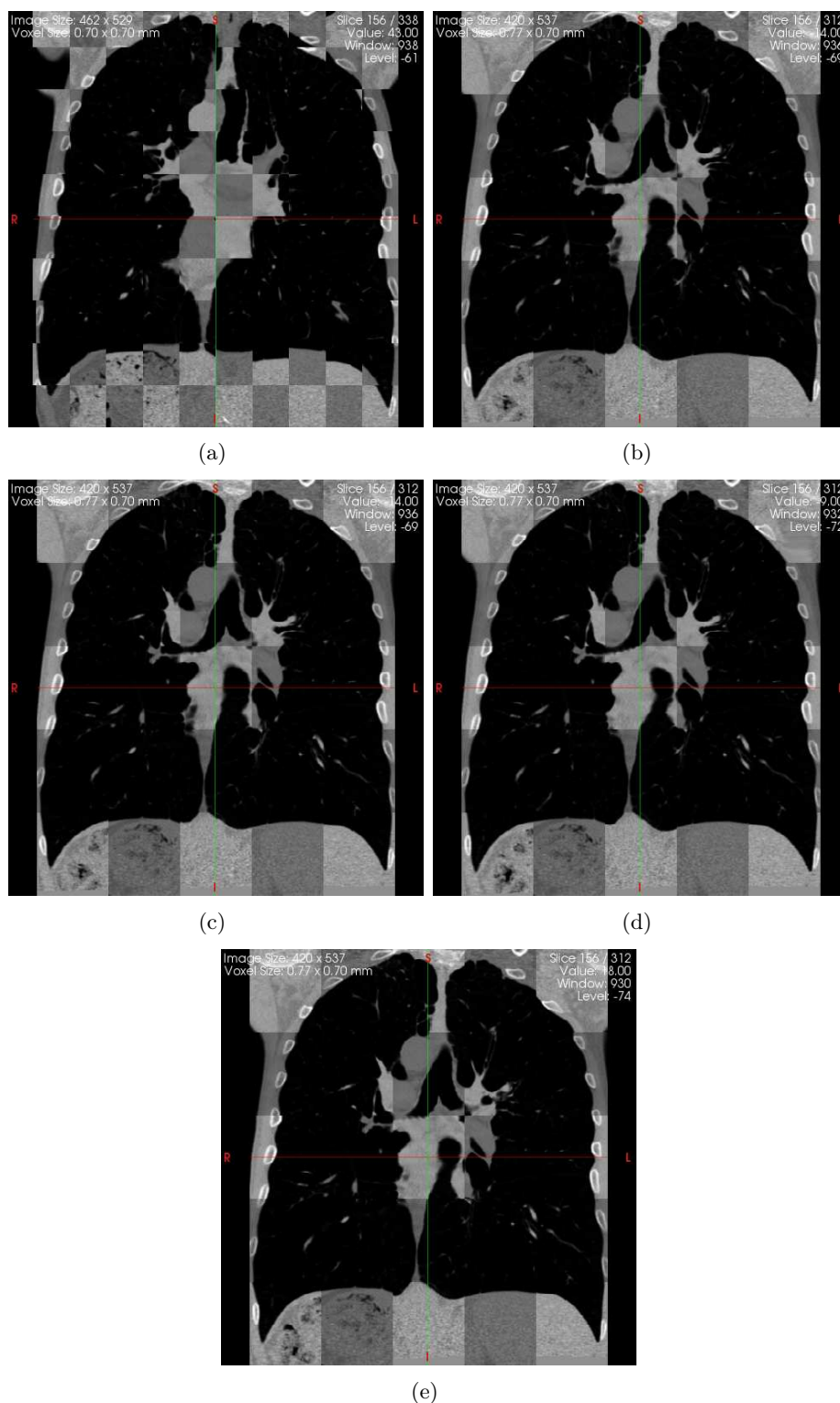


FIGURE 4.41 – Evaluation visuelle de la qualité des recalages en fonction du nombre d'itérations à chaque niveau de la pyramide, pour respectivement 50, 70, 150, et 200 itérations. La première image donne les images flottante et de référence, après le pré-recalage rigide.

Lung Boundary Results:

% Error means the percentage of points checked in the reference standard for which a penalty was incurred

Scan Pair	% Error Left Lung	% Error Right Lung	% Error Upper Lung	% Error Lower Lung	Score (% Error Overall)	Ranking
01	0.113738	0.179212	0.000336536	0.246395	0.145973	23
02	0.000998739	0.144289	0.0477258	0.088921	0.0699842	25
03	0.00504882	0.00342917	0.0034399	0.00503442	0.00427412	23
04	0.00962788	0	0.00792435	0	0.00399002	19
05	0	0	0	0	0	11
06	2.0733	2.19402	0.782565	3.192	2.13245	27
07	1.5832	4.80028	0.697365	4.93448	3.14211	27
08	0.0231564	0.026424	0	0.0447554	0.0247443	24
09	0.676027	0.205497	0.14591	0.706897	0.446254	27
10	0.00310043	0.000495581	0.00311684	5.48886e-05	0.00157862	10
11	0.463713	0.0300556	0.382112	0.14172	0.246363	26
12	0	0.471339	0.377358	0.100576	0.228128	27
13	0.534746	0.0064259	0.0245192	0.492448	0.271701	27
14	3.71474	4.68571	1.25725	6.69144	4.20896	27
15	0.000816831	0.0035871	0	0.00395596	0.00216391	23
16	0.0254413	0.0067457	0	0.0287328	0.0163137	23
17	0.0446265	0.00529136	0	0.0458262	0.0239068	26
18	0.426654	0.325677	0.00849067	0.67661	0.376835	25
19	0.759996	0	0.0389926	0.669317	0.398667	27
20	3.01675	2.85863	0.18679	5.16678	2.93815	27
21	0.633315	0.758445	0.00495681	1.23237	0.693977	23
22	0.104781	0.103353	0.0626247	0.142173	0.10409	26
23	0.00118273	0.131124	0.149557	0.00122735	0.0660707	25
24	0.00195164	0.00683583	0.000369045	0.00870289	0.00470661	23
25	0	0	0	0	0	11.5
26	0.076316	0.176705	0.285295	0.0133952	0.12785	27
27	0.124418	0.205113	0.209181	0.12741	0.164477	26
28	2.74621	3.84328	0.0190328	5.82698	3.29334	26
29	2.09934e-05	0.000178179	0.000225494	0	0.000108528	14
30	0.193638	3.07672	3.39509	0.0638814	1.5881	27
Avg	0.578584	0.808295	0.269674	1.02174	0.690842	23.4167

FIGURE 4.42 – Résultats quantitatifs de la base de données EMPIRE. Ce tableau donne les résultats du critère de l'alignement des frontières des poumons, pour les 30 patients de la base. En vert, les résultats parfaits au sens du critère. En jaune, les résultats très bons. En rouge, les résultats moyens.

Fissure Results:

% Error means the percentage of points checked in the reference standard for which a penalty was incurred

Scan Pair	% Error Left Lung	% Error Right Lung	Score (% Error Overall)	Ranking
01	4.57133	2.8643	3.19846	21
02	0	0	0	12
03	0.036036	0	0.00940855	25
04	0	0	0	13
05	0	0	0	13
06	0.781594	6.42112	4.21278	27
07	7.02665	4.92665	6.05028	23
08	4.06327	2.42893	3.09641	22
09	0.794948	0.0582635	0.318787	25
10	0	0	0	12.5
11	11.6038	1.73773	5.00014	23
12	0	0	0	10
13	1.4247	0.350483	1.04781	25
14	16.396	2.23242	8.82066	24
15	0	0.0836656	0.0457753	24
16	0.971136	0.0340977	0.259101	19
17	0.0923033	0.0135099	0.0440893	7
18	25.1397	0.568121	10.2649	25
19	0.14373	0	0.0637361	25
20	10.325	16.9719	14.4349	27
21	5.06382	5.55965	5.40124	22
22	0.627832	0	0.321451	22
23	3.22581	6.35546	5.10825	25
24	0	0	0	10
25	0	0	0	12
26	0.0101726	0.00901984	0.00935724	23
27	0.0950973	0.196231	0.155904	23
28	11.6345	13.1713	12.6219	24
29	0.00220699	0	0.000472545	22
30	0.00601181	0	0.00230774	22
Avg	3.46785	2.13276	2.68294	20.25

FIGURE 4.43 – Résultats quantitatifs de la base de données EMPIRE. Ce tableau donne les résultats du critère de l'alignement des scissures principales, pour les 30 patients de la base. En vert, les résultats parfaits au sens du critère. En jaune, les résultats très bons. En rouge, les résultats moyens.

Landmark Results:

All distances are measured in millimetres

Scan Pair	Avg Dist AP	Avg Dist SI	Avg Dist LR	Avg Dist Upper Lung	Avg Dist Lower Lung	Min Dist	Max Dist	Score (Avg Dist)	Ranking
01	3.72469	3.255	2.10023	6.50523	5.87772	0.679688	16.4888	6.1852	22
02	0.302344	0.343	0.309375	0.67336	0.797983	0	4.0865	0.733179	20
03	0.442969	0.546	0.464063	1.0571	0.999513	0	5.01039	1.02888	24
04	0.888526	1.21787	0.58069	2.60708	1.11415	0	8.96562	1.84947	17
05	0.0853125	0.035	0.091875	0.103799	0.275768	0	1.16247	0.188064	17
06	22.2422	10.56	5.52703	14.1511	33.5225	0.861786	70.9013	26.5488	27
07	3.2407	5.054	2.44406	6.88378	8.23466	0.773438	32.9908	7.49039	25
08	1.245	2.163	0.9525	2.65196	3.55442	0	14.3183	3.13928	22
09	0.676758	0.574	0.772461	1.03606	1.91218	0	12.1335	1.46536	25
10	2.26679	1.89844	1.02845	5.53809	1.75247	0.59998	18.8813	3.61703	19
11	1.56979	2.38	0.98918	2.49016	4.63357	0	20.6818	3.45122	25
12	0.341172	0.308	0.263633	0.478554	0.792621	0	7.55916	0.651291	21
13	0.742187	1.3125	0.644531	1.80535	2.13131	0	9.86086	1.98137	25
14	6.86672	8.547	4.12805	11.3545	15.6509	0	44.1074	13.4597	25
15	0.611719	0.525	0.442969	0.893381	1.33478	0	6.19251	1.14056	22
16	0.810547	1.325	0.703125	2.15368	2.00979	0	6.11537	2.07886	23
17	0.64453	0.92	0.468749	1.23486	2.0752	0	8.11833	1.59672	24
18	4.92871	5.614	2.87793	8.02507	10.3049	0.683594	23.9566	9.16497	25
19	0.64625	0.567	0.403906	0.901913	1.55803	0	16.9644	1.20372	25
20	6.02973	8.813	3.59746	12.7707	13.086	1.66458	36.6719	12.9031	24
21	4.92981	6.28572	3.17439	9.56248	10.7044	0	29.4437	10.0825	21
22	0.960938	0.896	0.851562	1.95062	1.75241	0	7.4682	1.83962	23
23	0.615234	1.325	0.507812	1.77206	2.20558	0	7.56331	2.00183	24
24	0.438343	0.831791	0.377308	1.29486	1.01314	0	3.30254	1.17641	14
25	0.170625	0.35	0.149297	0.531418	0.594829	0	3.09642	0.559953	23
26	0.73289	0.645	0.463242	1.8946	0.860346	0	11.0042	1.3361	24
27	0.43084	1.064	0.513984	1.14396	1.9084	0	9.57596	1.54147	24
28	2.6417	7.385	4.01092	10.3806	9.56734	0.744141	38.2731	9.8539	24
29	0.715389	1.08834	0.471248	1.85662	1.23931	0	5.11252	1.59103	17
30	0.592578	0.567	0.344336	1.38646	0.80566	0	19.061	1.09025	24
Avg	2.35117	2.54652	1.32181	3.83831	4.74233	0.20024	16.6356	4.36501	22.5

FIGURE 4.44 – Résultats quantitatifs de la base de données EMPIRE. Ce tableau donne les résultats du critère de la correspondance entre points annotés, pour les 30 patients de la base. En jaune, les résultats très bons. En rouge, les résultats moyens.

Singularity Results:

% Error means the percentage of points checked in the lung volume for which a penalty was incurred

Scan Pair	% Error Left Lung	% Error Right Lung	% Error Upper Lung	% Error Lower Lung	Score (% Error Overall)	Ranking
01	0	0	0	0	0	9.5
02	0	0	0	0	0	11
03	0	0	0	0	0	11.5
04	0	0	0	0	0	11
05	0	0	0	0	0	10.5
06	0	0	0	0	0	12.5
07	0	0	0	0	0	8.5
08	0	0	0	0	0	11
09	0	0	0	0	0	10.5
10	0	0	0	0	0	10.5
11	0	0	0	0	0	10
12	0	0	0	0	0	12
13	0	0	0	0	0	10.5
14	0	0	0	0	0	7
15	0	0	0	0	0	11
16	0	0	0	0	0	11.5
17	0	0	0	0	0	11
18	0	0	0	0	0	8.5
19	0	0	0	0	0	11
20	0	0	0	0	0	9.5
21	0	0	0	0	0	9
22	0	0	0	0	0	10.5
23	0	0	0	0	0	11
24	0	0	0	0	0	10.5
25	0	0	0	0	0	10
26	0	0	0	0	0	12
27	0	0	0	0	0	11.5
28	0	0	0	0	0	10
29	0	0	0	0	0	10.5
30	0	0	0	0	0	11.5
Avg	0	0	0	0	0	10.5

FIGURE 4.45 – Résultats quantitatifs de la base de données EMPIRE. Ce tableau donne les résultats du critère des singularités dans le champ de déformation dense, pour les 30 patients de la base. Tous les recalages sont parfaits au sens de ce critère.

Conclusion & Perspectives

Le recalage d'images médicales est un outil devenu presque indispensable au praticien, du fait du volume et de la diversité des images médicales produites aujourd'hui. Il permet une analyse automatique des images permettant de fusionner ou de comparer les informations provenant de modalités d'imagerie différentes. Dans ce contexte, les travaux réalisés durant cette thèse ont tenté d'apporter de nouveaux éléments de réponse au problème du recalage, qui reste délicat. Plus spécifiquement, nous nous sommes attachés à la problématique suivante : « *Approximation de l'Information Mutuelle basée sur le développement d'Edgeworth : application au recalage d'images médicales* ».

Les travaux réalisés dans cette thèse ont porté essentiellement sur la définition et la mise en oeuvre de nouvelles mesures de similarité pour le recalage rigide et non-rigide d'images médicales mono- et multimodales.

Une étude bibliographique nous a tout d'abord convaincu d'utiliser l'information mutuelle, qui malgré quelques limitations relatives à son estimation, se révèle particulièrement efficace dans le cadre du recalage d'images multimodales. Nous avons alors cherché à proposer une nouvelle méthode d'estimation qui tend à minimiser les défauts inhérents aux méthodes d'estimation classiques. Pour ce faire, nous nous sommes inspirés des travaux menés en traitement du signal, et plus particulièrement en analyse en composantes indépendantes. Dans ce contexte, des approximations de l'information mutuelle existent. Elles se basent sur l'utilisation des cumulants, statistiques qui permettent de caractériser une distribution de probabilité. Malheureusement, les approximations de l'information mutuelle utilisées en analyse en composantes indépendantes s'appuient sur l'hypothèse d'indépendance et notamment de décorrélation des variables aléatoires mises en jeu qui ne peuvent être utilisées dans notre contexte. Aussi avons nous décidé de développer nos propres mesures, qui ne font plus cette hypothèse.

Nous nous sommes alors orientés vers le développement d'Edgeworth, qui permet d'approximer une densité de probabilité à partir des cumulants et des polynômes d'Hermite des variables aléatoires considérées. La contribution principale de ce travail a été de proposer une approximation de l'entropie conjointe, pour des variables aléatoires potentiellement corrélées, se basant sur un tel développement. Nous avons également présenté une approximation de l'entropie marginale en utilisant les mêmes développements mathématiques. Cela nous a conduit à la définition de nouvelles approximations de l'information mutuelle et de l'information mutuelle normalisée, pour le recalage d'images médicales.

La phase de validation de ces nouvelles mesures a été importante. Nous avons comparé notre approche avec celles classiquement utilisées. Les premiers tests nous ont convaincu de l'efficacité de notre méthodologie. Mais ils ont également montré le comportement singulier de notre approximation basée sur un développement d'Edgeworth tronqué à l'ordre 4. Nous avons donc sélectionné, parmi les mesures que nous proposons, celle approximant le mieux l'information mutuelle. Après avoir incorporé cette mesure dans un schéma de recalage global

comportant une méthode d'optimisation efficace, et un modèle de déformation reconnu, nous avons montré que l'approche proposée permettait de recalcr efficacement des images médicales mono- et multimodales, dans un cadre rigide et non-rigide. La généralité de l'approche la rend candidate pour de nombreux problèmes nécessitant l'utilisation d'une méthode de recalage, et plus généralement d'une mesure de similarité.

Les perspectives offertes par ce travail sont nombreuses. A court terme tout d'abord, il va être intéressant de mettre à l'épreuve notre mesure de similarité basée sur un développement d'Edgeworth tronqué à l'ordre 4. Celle-ci, bien que singulière, montre des qualités de régularité qui en font une mesure probablement efficace dans notre contexte. Nous pourrions également envisager de construire une mesure à mi-chemin entre les approximations tronquées à l'ordre 3 et 4, en utilisant le développement de Gram-Charlier à la place de celui d'Edgeworth. Cela permettrait de se prémunir du terme du développement à l'ordre 4 qui rend les approximations singulières. Il nous faut également mener une évaluation de l'estimée de l'information mutuelle normalisée basée sur un développement tronqué à l'ordre 3 (3EIMN). Dans nos tests préliminaires de régularité, cette mesure montrait des résultats similaires à 3EIM que nous avons étudiée en détail. Il faudra donc montrer si l'apport de la normalisation dans nos mesures est aussi conséquent qu'avec les mesures traditionnelles.

Il apparaît également nécessaire de développer une dérivée analytique de nos mesures, qui autoriserait une optimisation encore plus efficace. En effet, la plupart des optimiseurs utilisés en recalage se basent sur le gradient de la fonction de coût pour converger vers l'optimum. Aussi, un gradient calculé analytiquement est assurément plus précis que l'estimation que nous utilisons pour le moment, et permet donc une meilleure convergence de l'optimiseur.

Enfin, toujours dans un avenir proche, il serait intéressant de réimplémenter nos mesures de similarité avec la structure de la librairie *elastiX*, en utilisant notamment les méthodes de la classe *AdvancedImageToImageMetric* qui permet d'améliorer le rendement de ces mesures. Cela peut paraître anecdotique, mais tout au long de ce travail, nous avons été confronté au problème des coûts calculatoires, surtout pour le recalage non-rigide. Ainsi, même un gain de temps infime sur une évaluation de la fonction de coût peut représenter un gain vraiment conséquent dans une procédure complète de recalage.

La confrontation de notre méthode à des cas cliniques réels est également à poursuivre, car nous avons vu que cela nous mène à de nouvelles hypothèses de travail. Parmi celles-ci, l'utilisation de mesures multidimensionnelles est certainement une des plus prometteuses, et sa mise en oeuvre avec notre mesure est possible sans que la charge calculatoire ne devienne trop importante. Cela va dans le sens des avancées récemment proposées par la communauté scientifique, et permettrait d'inclure d'autres informations que celles d'intensités dans nos mesures. Nous pensons notamment que prendre en compte l'information spatiale donnée par les images permettrait d'améliorer les performances de nos estimateurs.

Une autre idée à développer est de considérer notre approximation non plus au voisinage d'une gaussienne, mais plutôt d'une combinaison de gaussiennes. Pour ce faire, un pré-traitement est nécessaire pour déterminer les paramètres des différentes gaussiennes approximaant les distributions de probabilité des nos images. Cette méthode permettrait d'estimer plus fidèlement nos densités de probabilités, et donnerait en conséquence de meilleures estimées de l'information mutuelle.

D'un point de vue plus global, d'autres composantes des méthodes de recalage nécessitent, à nos yeux, d'être étudiées. Des travaux méritent notamment d'être menés sur les méthodes d'optimisation utilisées en recalage. L'introduction d'algorithmes à pas optimal par itération, par exemple, permettrait sans doute de réduire les coûts de calcul et de s'approcher plus efficacement de l'optimum. Considérer un pas différent pour chaque paramètre à optimiser,

contrairement au pas scalaire qui est toujours utilisé dans notre contexte, nous semble également envisageable. Cela permettrait de considérer plus efficacement les déformations locales intervenant dans les problèmes de recalage non-rigide.

Annexes

Annexe A

Transformations rigides et affines d'une image

Matrices de transformation 3D

Matrice de rotation suivant la direction x :

$$\mathbf{R}_x = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(\theta) & -\sin(\theta) & 0 \\ 0 & \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Matrice de rotation suivant la direction y :

$$\mathbf{R}_y = \begin{pmatrix} \cos(\theta) & 0 & -\sin(\theta) & 0 \\ 0 & 1 & 0 & 0 \\ \sin(\theta) & 0 & \cos(\theta) & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Matrice de rotation suivant la direction z :

$$\mathbf{R}_z = \begin{pmatrix} \cos(\theta) & -\sin(\theta) & 0 & 0 \\ \sin(\theta) & \cos(\theta) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Matrice de translation :

$$\mathbf{V} = \begin{pmatrix} 1 & 0 & 0 & t_x \\ 0 & 1 & 0 & t_y \\ 0 & 0 & 1 & t_z \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Matrice de facteur d'échelle (scaling) :

$$\mathbf{S}_c = \begin{pmatrix} sc_x & 0 & 0 & 0 \\ 0 & sc_y & 0 & 0 \\ 0 & 0 & sc_z & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Matrice de cisaillement (shearing) :

$$\mathbf{Sh} = \begin{pmatrix} 1 & sh_{xy} & sh_{xz} & 0 \\ sh_{yx} & 1 & sh_{yz} & 0 \\ sh_{zx} & sh_{zy} & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Le produit de ces matrices permet de décrire à la fois des matrices de déformation rigides et affines :

- Matrice de déformation rigide :

$$\mathbf{M_R} = \mathbf{R_x R_y R_z T}$$

- Matrice de déformation affine :

$$\mathbf{M_A} = \mathbf{R_x R_y R_z T S c Sh}$$

Pour transformer les voxels d'une image flottante de coordonnées (x_F, y_F, z_F) à partir des voxels d'une image de référence de coordonnées (x_R, y_R, z_R) , on écrit :

$$\begin{pmatrix} x_F \\ y_F \\ z_F \end{pmatrix} = \mathbf{M} \begin{pmatrix} x_R \\ y_R \\ z_R \end{pmatrix}$$

où \mathbf{M} est une matrice égale à $\mathbf{M_R}(1 : 3, :)$ ou $\mathbf{M_A}(1 : 3, :)$ selon le type de transformation recherchée (rigide ou affine), en prenant comme convention d'écriture que $\mathbf{B}(1 : 3, :)$ représente le bloc matriciel extrait de \mathbf{B} en prenant les 3 premières lignes de la matrice \mathbf{B} .

Annexe B

B-Splines

Les B-Splines sont des polynômes continus par morceaux qui ont à l'origine été utilisés pour le dessin assisté par ordinateur, où ils servent à décrire des courbes et surfaces 3D. Ces B-Splines sont désormais très populaires en recalage d'images médicales où elles servent de modèle de transformation pour décrire un champ dense de déformation. Elles y ont été introduites par [Rueckert et al., 1999] et leur popularité ne s'est jamais démentie, puisque récemment encore [Klein et al., 2009], elles restent une référence dans le domaine. Ceci est dû à leurs intéressantes propriétés. Nous rappelons dans cette annexe les plus importantes d'entre-elles, ou du moins celles qui nous ont été utiles au cours de cette thèse. Le lecteur pourra se référer à [Piegl and Tiller, 1997] pour de plus amples détails concernant les B-Splines et leur généralisation.

Les B-Splines peuvent intervenir à trois niveaux dans la procédure de recalage :

- pour le modèle de déformation tout d'abord, puisque elles permettent de modéliser des déformations non-rigides complexes.
- pour l'interpolation.
- pour l'estimation des densités de probabilités, où elles servent de noyau pour les estimateurs par fenêtrage de Parzen.

Certains auteurs [Thévenaz and Unser, 1997, Mattes et al., 2003] mettent en avant le fait qu'un modèle unifié, où les B-Splines sont utilisées pour les trois tâches présentées ci-dessus, est très efficace car il apporte des simplifications importantes.

Définition

Les B-Splines sont des fonctions symétriques en forme de cloche. Mathématiquement, la B-Spline de degré n est obtenue par n convolutions de la fonction rectangle $\beta^0(z) = \Pi(z, 1)$:

$$\Pi(z, \epsilon) = \begin{cases} \frac{1}{\epsilon} & \text{si } |z| < \frac{\epsilon}{2} \\ \frac{0.5}{\epsilon} & \text{si } |z| = \frac{\epsilon}{2} \\ 0 & \text{si } |z| > \frac{\epsilon}{2} \end{cases} \quad (\text{B.1})$$

avec ϵ la largeur du support.

On a donc :

$$\beta^{(n)}(z) = \underbrace{\left(\beta^{(0)} * \beta^{(0)} * \dots * \beta^{(0)} \right)}_{(n+1) \text{ termes}}(z) \quad (\text{B.2})$$

Les formules explicites des B-Splines de degré 0 à 3 sont données par :

$$\beta^{(0)}(z) = \begin{cases} 1 & \text{si } |z| < \frac{1}{2} \\ \frac{1}{2} & \text{si } |z| = \frac{1}{2} \\ 0 & \text{si } |z| > \frac{1}{2} \end{cases} \quad (\text{B.3})$$

$$\beta^{(1)}(z) = \begin{cases} 1 - |z| & \text{si } |z| < 1 \\ 0 & \text{si } |z| \geq 1 \end{cases}$$

$$\beta^{(2)}(z) = \begin{cases} \frac{3}{4} - |z|^2 & \text{si } |z| < \frac{1}{2} \\ \frac{1}{8}(2|z| - 3)^2 & \text{si } \frac{1}{2} \leq |z| < \frac{3}{2} \\ 0 & \text{si } |z| \geq \frac{3}{2} \end{cases}$$

$$\beta^{(3)}(z) = \begin{cases} \frac{2}{3} - |z|^2 + \frac{1}{2}|z|^3 & \text{si } |z| < 1 \\ -\frac{1}{6}(|z| - 2)^3 & \text{si } 1 \leq |z| < 2 \\ 0 & \text{si } |z| \geq 2 \end{cases}$$

Leur représentation est donnée figure B.1 jusqu'à l'ordre 3. Ces polynômes sont continus par morceaux, et chaque morceau de degré n est C^n continu. Au niveau de la transition entre 2 B-Splines, une continuité C^{n-1} est préservée.

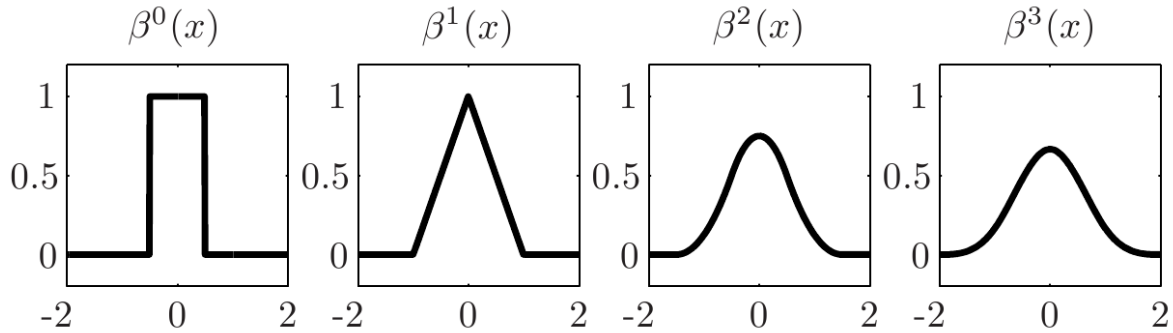


FIGURE B.1 – Représentation des B-Splines d'ordre 0 à 3.

Propriétés

Les propriétés importantes des B-Splines sont :

le calcul explicite de la dérivée

Cette propriété est utilisée pour calculer les dérivées du champ de transformation pendant l'optimisation :

$$\frac{d\beta^{(n)}(z)}{dz} = \beta^{(n-1)}\left(z + \frac{1}{2}\right) - \beta^{(n-1)}\left(z - \frac{1}{2}\right) \quad (\text{B.4})$$

la partition de l'unité

Cette propriété permet d'utiliser le noyau B-Spline pour la construction des histogrammes conjoints, par fenêtrage de Parzen :

$$\forall z \in \mathbb{R}, \sum_{k=-\infty}^{+\infty} \beta^{(n)}(z+k) = 1 \quad (\text{B.5})$$

Cette propriété assure la contribution unitaire d'une paire de pixel dans la construction de l'histogramme conjoint.

un support limité

Cela permet un calcul efficace des champs de déplacement, étant donné que seuls les vecteurs de déplacement dans le voisinage du PDC sont pris en compte. Cette propriété s'écrit :

$$\beta^{(n)}(z) \neq 0 \Leftrightarrow z \in \left] \frac{-n-1}{2}, \frac{n+1}{2} \right[\quad (\text{B.6})$$

Cette dernière propriété peut sembler assez restrictive. Cependant, il est possible d'obtenir un support plus large, en étendant les B-Splines par un facteur ϵ :

$$\beta_{\epsilon}^{(n)}(z) = \frac{\beta^{(n)}(\frac{z}{\epsilon})}{\epsilon}, \quad \frac{\partial \beta_{\epsilon}^{(n)}(z)}{\partial z} = \beta_{\epsilon}^{(n)'}(z) = \frac{\beta^{(n)'}(\frac{z}{\epsilon})}{\epsilon^2} \quad (\text{B.7})$$

Pour la construction des histogrammes conjoints, cette propriété est assez intéressante : elle permet d'étendre la contribution d'un pixel à un voisinage plus important, et donc de lisser davantage l'histogramme. Il y a clairement un compromis à faire entre le degré de la spline à prendre en compte, et la largeur de son support.

Bibliographie

- [Klein et al., 2009] Klein, A., Andersson, J., Ardekani, B., Ashburner, J., Avants, B., Chiang, M.-C., Christensen, G., Collins, D. L., Hellier, P., Song, J. H., Jenkinson, M., Lepage, C., Rueckert, D., Thompson, P., Vercauteren, T., Woods, R., Mann, J., and Parsey, R. (2009). Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. *NeuroImage*, 46(3) :786–802.
- [Mattes et al., 2003] Mattes, D., Haynor, D. R., Vesselle, H., Lewellen, T. K., and Eubank, W. (2003). PET-CT image registration in the chest using Free-Form Deformations. *IEEE Transactions on Medical Imaging*, 22(1) :120–128.
- [Piegl and Tiller, 1997] Piegl, L. and Tiller, W. (1997). *The NURBS book (2nd ed.)*. Springer-Verlag New York, Inc., New York, NY, USA.
- [Rueckert et al., 1999] Rueckert, D., Sonoda, L., Hayes, C., Hill, D., Leach, M., and Hawkes, D. (1999). Nonrigid registration using free-form deformations : Application to breast MR images. *IEEE Transactions on Medical Imaging*, 18(8) :712–721.
- [Thévenaz and Unser, 1997] Thévenaz, P. and Unser, M. (1997). Spline pyramids for inter-modal image registration using mutual information. In *Proceedings of the SPIE Conference on Mathematical Imaging : Wavelet Applications in Signal and Image Processing V*, volume 3169, pages 236–247, San Diego CA, USA.

Annexe C

Développements mathématiques

Dans cette partie, nous présentons nos approximations des entropies marginales et conjointes par un développement d'Edgeworth à l'ordre 4, sous l'hypothèse que les variables ont préalablement été centrées-réduites. Ces développements sont l'apport principal de ce travail de thèse.

Nous faisons dans toute cette annexe l'hypothèse que les données sont centrées-réduites, c'est à dire de moyenne nulles et de variance unité. Et c'est également le cas des distributions gaussiennes utilisées, puisque les développements d'Edgeworth et de Gram-Charlier cherchent à approximer une densité de probabilité au voisinage d'une gaussienne ayant les mêmes statistiques de premier et second ordre, c'est à dire les mêmes moyennes et variances.

C.1 Approximation de l'entropie marginale par un développement d'Edgeworth

Nous partons de l'équation du développement univarié d'Edgeworth autour de la distribution normale jusqu'à l'ordre 4 de la densité de probabilité marginale :

$$\begin{aligned} p_{x^{(i)}}(u^{(i)}) &= \phi_{x^{(i)}}(u^{(i)}) \left(1 + \frac{1}{3!} \kappa_3 H_3(u^{(i)}) \right. \\ &\quad \left. + \frac{1}{4!} \kappa_4 H_4(u^{(i)}) + \frac{10}{6!} \kappa_3^2 H_6(u^{(i)}) \right) + O(N^{-3/2}) \end{aligned} \quad (\text{C.1})$$

et nous considérons implicite la dépendance envers la VA (i) afin d'alléger les notations. (C.1) se réécrit donc :

$$\begin{aligned} p_x(u) &= \phi_x(u) \left(1 + \frac{1}{3!} \kappa_3 H_3(u) \right. \\ &\quad \left. + \frac{1}{4!} \kappa_4 H_4(u) + \frac{10}{6!} \kappa_3^2 H_6(u) \right) + O(N^{-3/2}) \end{aligned} \quad (\text{C.2})$$

L'idée est d'exprimer l'entropie de Shannon ou entropie différentielle en fonction de cette estimation de la densité de probabilité. Nous rappelons la formule de cette entropie :

$$H(x) = - \int_{\mathbb{R}} p_x(u) \log(p_x(u)) du \quad (\text{C.3})$$

Pour le moment, pour simplifier encore notre raisonnement, nous posons :

$$A_x(u) = \frac{1}{3!} \kappa_3 H_3(u) + \frac{1}{4!} \kappa_4 H_4(u) + \frac{10}{6!} \kappa_3^2 H_6(u)$$

L'entropie de Shannon (C.3) peut alors s'exprimer en fonction de (C.2) par :

$$H(x) \approx - \int_{\mathbb{R}} \phi_x(u) (1 + A_x(u)) \log [\phi_x(u) (1 + A_x(u))] du \quad (C.4)$$

$$\approx - \int_{\mathbb{R}} \phi_x(u) (1 + A_x(u)) \log (\phi_x(u)) du \quad (C.5)$$

$$- \int_{\mathbb{R}} \phi_x(u) (1 + A_x(u)) \log (1 + A_x(u)) du \quad (C.6)$$

Nous allons simplifier (C.6) en utilisant le développement de Taylor $(1+x) \log(1+x) \approx x + \frac{x^2}{2}$ autour de zéro [Comon, 1994, Van Hulle, 2005]. C'est la seule simplification nécessaire pour mener à bien le développement. Il s'en suit que (C.4) peut se réécrire :

$$H(x) \approx - \int_{\mathbb{R}} \phi_x(u) \log (\phi_x(u)) du \quad (C.7)$$

$$- \int_{\mathbb{R}} \phi_x(u) \log (\phi_x(u)) A_x(u) du \quad (C.8)$$

$$- \int_{\mathbb{R}} \phi_x(u) \left(A_x(u) + \frac{(A_x(u))^2}{2} \right) du \quad (C.9)$$

Nous allons maintenant résoudre explicitement les 3 termes (C.7), (C.8) et (C.9) de cette équation. Tout d'abord on reconnaît dans (C.7) l'entropie d'un loi normale et on a donc :

$$- \int_{\mathbb{R}} \phi_x(u) \log (\phi_x(u)) du = \log(\sqrt{2\pi e}) = 0.5 \log(2\pi e) \quad (C.10)$$

Nous devons ensuite résoudre (C.8). Pour cela, nous commençons par réécrire la définition d'une distribution normale monovariée d'espérance $E[x]$ et de variance $\text{Var}(x)$:

$$\phi_x(u) = \frac{1}{\sqrt{2\pi \text{Var}(x)}} e^{-(u-E[x])^2/(2\text{Var}(x))} \quad (C.11)$$

Sous l'hypothèse de données centrées-réduites, cette équation (C.11) se réécrit :

$$\phi_x(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} \quad (C.12)$$

Nous pouvons donc écrire :

$$\log \phi_x(u) = \log \left(\frac{1}{\sqrt{2\pi}} \right) - \frac{u^2}{2} \quad (C.13)$$

et (C.8) s'exprime alors par :

$$\begin{aligned} - \int_{\mathbb{R}} \phi_x(u) \log (\phi_x(u)) A_x(u) du &= - \int_{\mathbb{R}} \phi_x(u) A_x(u) \left(\log \left(\frac{1}{\sqrt{2\pi}} \right) - \frac{u^2}{2} \right) du \\ &= - \log \left(\frac{1}{\sqrt{2\pi}} \right) \int_{\mathbb{R}} \phi_x(u) A_x(u) du \end{aligned} \quad (C.14)$$

$$+ \frac{1}{2} \int_{\mathbb{R}} \phi_x(u) A_x(u) u^2 du \quad (C.15)$$

Nous allons maintenant réintroduire la propriété d'orthogonalité des polynômes d'Hermite présentée au chapitre 3, qui s'écrit :

$$\int_{\mathbb{R}} \phi_x(u) H_i(u) H_j(u) du = \begin{cases} i! & \text{si } i = j \\ 0 & \text{autrement} \end{cases} \quad (\text{C.16})$$

qui va nous être utile tout au long de cette démonstration. Nous allons maintenant développer (C.14). En utilisant le fait que le premier polynôme d'Hermite $H_0(u)$ est égal à 1, nous pouvons écrire :

$$\begin{aligned} & - \log \left(\frac{1}{\sqrt{2\pi}} \right) \int_{\mathbb{R}} \phi_x(u) A_x(u) du \\ &= - \log \left(\frac{1}{\sqrt{2\pi}} \right) \int_{\mathbb{R}} \phi_x(u) \left(\frac{1}{3!} \kappa_3 H_3(u) + \frac{1}{4!} \kappa_4 H_4(u) + \frac{10}{6!} \kappa_3^2 H_6(u) \right) du \\ &= - \log \left(\frac{1}{\sqrt{2\pi}} \right) \int_{\mathbb{R}} \phi_x(u) \frac{1}{3!} \kappa_3 H_3(u) H_0(u) du \end{aligned} \quad (\text{C.17})$$

$$- \log \left(\frac{1}{\sqrt{2\pi}} \right) \int_{\mathbb{R}} \phi_x(u) \frac{1}{4!} \kappa_4 H_4(u) H_0(u) du \quad (\text{C.18})$$

$$- \log \left(\frac{1}{\sqrt{2\pi}} \right) \int_{\mathbb{R}} \phi_x(u) \frac{10}{6!} \kappa_3^2 H_6(u) H_0(u) du \quad (\text{C.19})$$

Les termes (C.17), (C.18) et (C.19) vont être nuls par la propriété d'orthogonalité des polynômes d'Hermite (C.16). En effet, pour (C.17) par exemple, nous pouvons écrire :

$$- \log \left(\frac{1}{\sqrt{2\pi}} \right) \int_{\mathbb{R}} \phi_x(u) \frac{1}{3!} \kappa_3 H_3(u) H_0(u) du = C \int_{\mathbb{R}} \phi_x(u) H_3(u) H_0(u) du = 0 \quad (\text{C.20})$$

où C représente une constante par rapport à la variable d'intégration. Il en va de même pour (C.18) et (C.19). Ainsi la première partie (C.14) de (C.8) est nulle. Il reste à résoudre la deuxième partie (C.15) de cette même expression. Pour cela, nous allons utiliser le polynôme d'Hermite de degré 2 défini par $H_2(u) = u^2 - 1 \Leftrightarrow H_2(u) + 1 = u^2$. Nous pouvons réécrire (C.15) comme :

$$\frac{1}{2} \int_{\mathbb{R}} \phi_x(u) A_x(u) u^2 du = \frac{1}{2} \int_{\mathbb{R}} \phi_x(u) A_x(u) H_2(u) du + \frac{1}{2} \int_{\mathbb{R}} \phi_x(u) A_x(u) du \quad (\text{C.21})$$

Encore une fois, ces deux intégrales vont être nulles par propriété d'orthogonalité des polynômes d'Hermite. La première fait intervenir des termes de type $C \int_{\mathbb{R}} \phi_x(u) H_j(u) H_2(u) du$ avec $j \neq 2$. Quand à la seconde, nous pouvons encore une fois la multiplier par $H_0(u) = 1$. Ainsi, (C.14) et (C.15), et par conséquent (C.8) est nulle. Dans notre définition de départ de l'entropie de Shannon approximée par un développement d'Edgeworth (C.4), il ne nous reste que le terme (C.9) à résoudre. Celui-ci s'exprime par :

$$- \int_{\mathbb{R}} \phi_x(u) \left(A_x(u) + \frac{(A_x(u))^2}{2} \right) du = - \int_{\mathbb{R}} \phi_x(u) A_x(u) du - \int_{\mathbb{R}} \phi_x(u) \frac{(A_x(u))^2}{2} du \quad (\text{C.22})$$

La première intégrale est nulle en faisant appel à $H_0(u)$. Il reste à développer la seconde intégrale :

$$\begin{aligned}
& - \int_{\mathbb{R}} \phi_x(u) \frac{(A_x(u))^2}{2} du \\
& = -\frac{1}{2} \int_{\mathbb{R}} \phi_x(u) \left(\frac{1}{3!} \kappa_3 H_3(u) + \frac{1}{4!} \kappa_4 H_4(u) + \frac{10}{6!} \kappa_3^2 H_6(u) \right)^2 du \\
& = -\frac{1}{2} \int_{\mathbb{R}} \phi_x(u) \left(\frac{1}{(3!)^2} \kappa_3^2 (H_3(u))^2 + \frac{1}{(4!)^2} \kappa_4^2 (H_4(u))^2 \right. \\
& + \frac{100}{(6!)^2} \kappa_3^4 (H_6(u))^2 + \frac{20}{3!6!} \kappa_3^3 H_3(u) H_6(u) \\
& \left. + \frac{2}{3!4!} \kappa_3 \kappa_4 H_3(u) H_4(u) + \frac{20}{4!6!} \kappa_3^2 \kappa_4 H_4(u) H_6(u) \right) du
\end{aligned} \tag{C.23}$$

Pour simplifier (C.23), nous utilisons les identités suivantes :

$$\int_{\mathbb{R}} u^{2k+1} \phi(u) du = 0 \tag{C.24}$$

et :

$$\int_{\mathbb{R}} u^{2k} \phi(u) du = 1.3...(2k-1) \tag{C.25}$$

qui découlent de la définition des moments d'une loi normale centrée-réduite. Nous allons développer les différents termes de (C.23) en fonction des formules explicites des polynômes d'Hermite ($H_3(u) = u^3 - 3u$, $H_4(u) = u^4 - 6u^2 + 3$, $H_6(u) = u^6 - 15u^4 + 45u^2 - 15$) pour se ramener à des expressions de type (C.24) et (C.25). Si l'on considère la première intégrale du développement de (C.23), celle-ci s'écrit :

$$-\frac{1}{2} \int_{\mathbb{R}} \phi_x(u) \frac{1}{(3!)^2} \kappa_3^2 (H_3(u))^2 du = -\frac{1}{2.(3!)^2} \kappa_3^2 \int_{\mathbb{R}} \phi_x(u) (H_3(u))^2 du \tag{C.26}$$

avec $(H_3(u))^2 = (u^3 - 3u)^2 = u^6 - 6u^4 + 9u^2$. D'après (C.25), on peut écrire :

$$\begin{aligned}
\int_{\mathbb{R}} \phi_x(u) (H_3(u))^2 du &= \int_{\mathbb{R}} \phi_x(u) (u^6 - 6u^4 + 9u^2) du \\
&= \int_{\mathbb{R}} \phi_x(u) u^6 du - \int_{\mathbb{R}} \phi_x(u) 6u^4 du + \int_{\mathbb{R}} \phi_x(u) 9u^2 du \\
&= 5 \times 3 - 6 \times 3 + 9 = 6
\end{aligned} \tag{C.27}$$

Donc (C.26) s'écrit finalement :

$$-\frac{1}{2} \int_{\mathbb{R}} \phi_x(u) \frac{1}{(3!)^2} \kappa_3^2 (H_3(u))^2 du = -\frac{6}{2.(3!)^2} \kappa_3^2 = -\frac{1}{2} \frac{\kappa_3^2}{3!} \tag{C.28}$$

Suivant ce principe, nous pouvons résoudre toutes les intégrales de (C.23). Tous les résultats de nos calculs découlant de la propriété d'orthogonalité des polynômes d'Hermite sont donnés en annexe D. A l'aide de ces résultats, les différentes intégrales de (C.23) s'écrivent :

$$\begin{aligned}
& - \frac{1}{2} \int_{\mathbb{R}} \phi_x(u) \frac{1}{(4!)^2} \kappa_4^2 (H_4(u))^2 du = -\frac{1}{2} \frac{\kappa_4^2}{4!} \\
& - \frac{1}{2} \int_{\mathbb{R}} \phi_x(u) \frac{100}{(6!)^2} \kappa_3^4 (H_6(u))^2 du = -\frac{5}{72} (\kappa_3)^4 \\
& - \frac{1}{2} \int_{\mathbb{R}} \phi_x(u) \frac{20}{3!6!} \kappa_3^3 H_3(u) H_6(u) du = 0 \\
& - \frac{1}{2} \int_{\mathbb{R}} \phi_x(u) \frac{2}{3!4!} \kappa_3 \kappa_4 H_3(u) H_4(u) du = 0 \\
& - \frac{1}{2} \int_{\mathbb{R}} \phi_x(u) \frac{20}{4!6!} \kappa_3^2 \kappa_4 H_4(u) H_6(u) du = 0
\end{aligned} \tag{C.29}$$

Nous pouvons donc finalement écrire (C.23) comme :

$$- \int_{\mathbb{R}} \phi_x(u) \frac{(A_x(u))^2}{2} du = -\frac{1}{2} \frac{\kappa_3^2}{3!} - \frac{1}{2} \frac{\kappa_4^2}{4!} - \frac{5}{72} \kappa_3^4 \tag{C.30}$$

Dans l'équation (C.4), le dernier terme (C.9) a été résolu explicitement. En reprenant les résultats des termes (C.7), (C.8) et (C.9) que nous avons résolus, nous pouvons finalement écrire :

$$H(x) \approx 0.5 \log(2\pi e) - \frac{1}{2} \frac{\kappa_3^2}{3!} - \frac{1}{2} \frac{\kappa_4^2}{4!} - \frac{5}{72} \kappa_3^4 \tag{C.31}$$

l'approximation de l'entropie marginale de la VA x par un développement d'Edgeworth à l'ordre 4. Si l'on se contente du développement à l'ordre 3, cette approximation est donnée par :

$$H(x) \approx 0.5 \log(2\pi e) - \frac{1}{2} \frac{\kappa_3^2}{3!} \tag{C.32}$$

C.2 Approximation de l'entropie conjointe par un développement d'Edgeworth

C.2.1 Démonstration à l'ordre 3

La démarche est similaire à celle employée pour l'approximation de l'entropie marginale. Mais cette fois-ci, nous considérons l'approximation de la densité de probabilité conjointe $p_{\mathbf{x}}$ par un développement d'Edgeworth. Nous présentons pour l'instant le calcul à l'ordre 3, et nous verrons par la suite que l'extension à l'ordre 4 est instantanée. L'approximation de la densité de probabilité conjointe par un développement d'Edgeworth à l'ordre 3 est donnée par :

$$p_{\mathbf{x}}(\mathbf{u}) = \phi_{\mathbf{x}}(\mathbf{u}) \left(1 + \frac{1}{3!} \kappa^{i_1, i_2, i_3} H_{i_1, i_2, i_3}(\mathbf{u}) \right) O(N^{-1}) \tag{C.33}$$

où il faut considérer toutes les permutations possibles des indices i_1, i_2, i_3 . Dans le cas bivarié, ces indices peuvent prendre la valeur 1 ou 2. Par conséquent, les 8 termes de la somme $\frac{1}{3!} \kappa^{i_1, i_2, i_3} H_{i_1, i_2, i_3}(\mathbf{u})$, que nous notons $B_{\mathbf{x}}(\mathbf{u})$ pour faciliter la lecture, sont donc :

$$B_{\mathbf{x}}(\mathbf{u}) = \frac{1}{3!} (\kappa^{1,1,1} H_{111}(\mathbf{u}) + 3\kappa^{1,1,2} H_{112}(\mathbf{u}) + 3\kappa^{1,2,2} H_{122}(\mathbf{u}) + \kappa^{2,2,2} H_{222}(\mathbf{u})) \quad (\text{C.34})$$

Nous voulons approximer l'entropie conjointe de Shannon [Collignon et al., 1995], [Studholme and Hawkes, 1995] donnée par :

$$H(\mathbf{x}) = - \oint_{\mathbb{R}^2} p_{\mathbf{x}}(\mathbf{u}) \log(p_{\mathbf{x}}(\mathbf{u})) d\mathbf{u} \quad (\text{C.35})$$

qui peut s'exprimer en fonction de (C.33) par :

$$H(\mathbf{x}) \approx - \oint_{\mathbb{R}^2} \phi_{\mathbf{x}}(\mathbf{u}) (1 + B_{\mathbf{x}}(\mathbf{u})) \log [\phi_{\mathbf{x}}(\mathbf{u}) (1 + B_{\mathbf{x}}(\mathbf{u}))] d\mathbf{u} \quad (\text{C.36})$$

$$\begin{aligned} &\approx - \oint_{\mathbb{R}^2} \phi_{\mathbf{x}}(\mathbf{u}) (1 + B_{\mathbf{x}}(\mathbf{u})) \log (\phi_{\mathbf{x}}(\mathbf{u})) d\mathbf{u} \\ &- \oint_{\mathbb{R}^2} \phi_{\mathbf{x}}(\mathbf{u}) (1 + B_{\mathbf{x}}(\mathbf{u})) \log (1 + B_{\mathbf{x}}(\mathbf{u})) d\mathbf{u} \\ &\approx - \oint_{\mathbb{R}^2} \phi_{\mathbf{x}}(\mathbf{u}) \log (\phi_{\mathbf{x}}(\mathbf{u})) d\mathbf{u} \end{aligned} \quad (\text{C.37})$$

$$- \oint_{\mathbb{R}^2} \phi_{\mathbf{x}}(\mathbf{u}) \log (\phi_{\mathbf{x}}(\mathbf{u})) B_{\mathbf{x}}(\mathbf{u}) d\mathbf{u} \quad (\text{C.38})$$

$$- \oint_{\mathbb{R}^2} \phi_{\mathbf{x}}(\mathbf{u}) \left(B_{\mathbf{x}}(\mathbf{u}) + \frac{(B_{\mathbf{x}}(\mathbf{u}))^2}{2} \right) d\mathbf{u} \quad (\text{C.39})$$

en utilisant encore une fois le développement de Taylor $(1+x) \log(1+x) \approx x + \frac{x^2}{2}$ autour de zéro. Dans cette expression (C.36), la première (C.37) intégrale est l'entropie d'une loi normale multidimensionnelle qui est donnée par :

$$- \oint_{\mathbb{R}^2} \phi_{\mathbf{x}}(\mathbf{u}) \log (\phi_{\mathbf{x}}(\mathbf{u})) d\mathbf{u} = \log(\sqrt{(2\pi e)^2 |\boldsymbol{\kappa}_{\mathbf{x}}|}) = 1 + \log(2\pi) + 0.5 \log(1 - \rho^2) \quad (\text{C.40})$$

avec $|\boldsymbol{\kappa}_{\mathbf{x}}|$ le déterminant de la matrice de covariance présentée au chapitre 3. Les 2 autres intégrales (C.38) et (C.39) vont se simplifier grâce notamment aux propriétés d'orthogonalité des polynômes d'Hermite multivariés que nous rappelons :

$$\int H_{i_1, \dots, i_k}(\mathbf{u}) H_{j_1, \dots, j_\ell}(\mathbf{u}) \phi_{\mathbf{x}}(\mathbf{u}) d\mathbf{u} = \begin{cases} [k!] \kappa_{i_1, j_1}^{\mathbf{x}} \dots \kappa_{i_k, j_k}^{\mathbf{x}} & \text{si } \ell = k \\ 0 & \text{autrement} \end{cases} \quad (\text{C.41})$$

Développons ces intégrales en commençant par (C.38). Pour cela, nous commençons par nous servir de la définition de la loi normale bivariable centrée-réduite. En posant $\mathbf{u} = (u_1, u_2)^\top$, on peut écrire :

$$\phi_{\mathbf{x}}(\mathbf{u}) = \frac{1}{(2\pi) |\boldsymbol{\kappa}_{\mathbf{x}}|^{1/2}} e^{-\frac{1}{2}(\mathbf{u}^\top \boldsymbol{\kappa}_{\mathbf{x}}^{-1} \mathbf{u})} \quad (\text{C.42})$$

avec $\boldsymbol{\kappa}_{\mathbf{x}} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$, $|\boldsymbol{\kappa}_{\mathbf{x}}| = 1 - \rho^2$ et $\boldsymbol{\kappa}_{\mathbf{x}}^{-1} = \frac{1}{1 - \rho^2} \begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix}$. Par conséquent, (C.42) se réécrit :

$$\phi_{\mathbf{x}}(\mathbf{u}) = \frac{1}{(2\pi) \sqrt{1 - \rho^2}} \exp \left(-\frac{1}{2(1 - \rho^2)} (u_1^2 + u_2^2 - 2\rho u_1 u_2) \right) \quad (\text{C.43})$$

Le logarithme de cette expression est ainsi :

$$\log(\phi_{\mathbf{x}}(\mathbf{u})) = \log\left(\frac{1}{(2\pi)\sqrt{1-\rho^2}}\right) - \frac{1}{2(1-\rho^2)}(u_1^2 + u_2^2 - 2\rho u_1 u_2) \quad (\text{C.44})$$

A partir de cette expression, on peut réécrire la seconde intégrale (C.38) de (C.36) comme étant :

$$\begin{aligned} & - \oint_{\mathbb{R}^2} \phi_{\mathbf{x}}(\mathbf{u}) \log(\phi_{\mathbf{x}}(\mathbf{u})) B_{\mathbf{x}}(\mathbf{u}) d\mathbf{u} \\ &= - \oint_{\mathbb{R}^2} \phi_{\mathbf{x}}(\mathbf{u}) B_{\mathbf{x}}(\mathbf{u}) \left(\log\left(\frac{1}{(2\pi)\sqrt{1-\rho^2}}\right) - \frac{1}{2(1-\rho^2)}(u_1^2 + u_2^2 - 2\rho u_1 u_2) \right) d\mathbf{u} \\ &= - \log\left(\frac{1}{(2\pi)\sqrt{1-\rho^2}}\right) \oint_{\mathbb{R}^2} \phi_{\mathbf{x}}(\mathbf{u}) B_{\mathbf{x}}(\mathbf{u}) d\mathbf{u} \end{aligned} \quad (\text{C.45})$$

$$+ \frac{1}{2(1-\rho^2)} \oint_{\mathbb{R}^2} \phi_{\mathbf{x}}(\mathbf{u}) B_{\mathbf{x}}(\mathbf{u}) (u_1^2 + u_2^2 - 2\rho u_1 u_2) d\mathbf{u} \quad (\text{C.46})$$

Ici encore, nous avons 2 intégrales. La première (C.45) est nulle par propriété d'orthogonalité des tenseurs d'Hermite. En effet, nous pouvons :

$$\begin{aligned} & - \log\left(\frac{1}{(2\pi)\sqrt{1-\rho^2}}\right) \oint_{\mathbb{R}^2} \phi_{\mathbf{x}}(\mathbf{u}) B_{\mathbf{x}}(\mathbf{u}) d\mathbf{u} \\ &= - \log\left(\frac{1}{(2\pi)\sqrt{1-\rho^2}}\right) \oint_{\mathbb{R}^2} \phi_{\mathbf{x}}(\mathbf{u}) B_{\mathbf{x}}(\mathbf{u}) H_{00}(\mathbf{u}) d\mathbf{u} \end{aligned}$$

puisque l'on a $H_0(\mathbf{u}) = H_{00}(\mathbf{u}) = H_{000}(\mathbf{u}) = \dots = 1$ d'après la définition des tenseurs d'Hermite (3.10) donnée au chapitre 3. Ainsi

$$\begin{aligned} & \log\left(\frac{1}{(2\pi)\sqrt{1-\rho^2}}\right) \oint_{\mathbb{R}^2} \phi_{\mathbf{x}}(\mathbf{u}) B_{\mathbf{x}}(\mathbf{u}) H_{00}(\mathbf{u}) d\mathbf{u} \\ &= \frac{1}{3!} \log\left(\frac{1}{(2\pi)\sqrt{1-\rho^2}}\right) \oint_{\mathbb{R}^2} \phi_{\mathbf{x}}(\mathbf{u}) H_{00}(\mathbf{u}) \\ & \quad (\kappa^{1,1,1} H_{111}(\mathbf{u}) + 3\kappa^{1,1,2} H_{112}(\mathbf{u}) + 3\kappa^{1,2,2} H_{122}(\mathbf{u}) + \kappa^{2,2,2} H_{222}(\mathbf{u})) d\mathbf{u} \end{aligned} \quad (\text{C.47})$$

Les 4 intégrales découlant de cette dernière (C.47) vont être de la forme :

$$C \oint_{\mathbb{R}^2} \phi_{\mathbf{x}}(\mathbf{u}) H_{00}(\mathbf{u}) H_{ijk}(\mathbf{u}) d\mathbf{u} = 0 \quad (\text{C.48})$$

par propriété d'orthogonalité des tenseurs d'Hermite (on a $l \neq k$ dans (C.41)). La première intégrale (C.45) de (C.38) est donc nulle. Intéressons nous maintenant à la seconde intégrale (C.46) donnée par :

$$\frac{1}{2(1-\rho^2)} \oint_{\mathbb{R}^2} \phi_{\mathbf{x}}(\mathbf{u}) B_{\mathbf{x}}(\mathbf{u}) (u_1^2 + u_2^2 - 2\rho u_1 u_2) d\mathbf{u} \quad (\text{C.49})$$

En nous servant de la définition des tenseurs d'Hermite covariants (3.13), nous pouvons écrire :

$$H_{11}(\mathbf{u}) + H_{22}(\mathbf{u}) - 2\rho H_{12}(\mathbf{u}) = u_1^2 - \kappa_{1,1} + u_2^2 - \kappa_{2,2} - 2\rho(u_1 u_2 - \kappa_{1,2})$$

soit :

$$u_1^2 + u_2^2 - 2\rho u_1 u_2 = H_{11}(\mathbf{u}) + H_{22}(\mathbf{u}) - 2\rho H_{12}(\mathbf{u}) + \kappa_{1,1} + \kappa_{2,2} - 2\rho\kappa_{1,2}$$

avec κ_{i_1, i_2} les éléments de la matrice $\boldsymbol{\kappa}_x^{-1}$ qui sont des constantes par rapport à la variable d'intégration \mathbf{u} . Nous pouvons ainsi réécrire (C.49) comme étant :

$$\frac{1}{2(1-\rho^2)} \oint_{\mathbb{R}^2} \phi_x(\mathbf{u}) B_x(\mathbf{u}) (H_{11}(\mathbf{u}) + H_{22}(\mathbf{u}) - 2\rho H_{12}(\mathbf{u}) + C) d\mathbf{u}$$

Encore une fois ces intégrales vont être nulles grâce à la propriété d'orthogonalité des polynômes d'Hermite puisque $B_x(\mathbf{u})$ contient des polynômes d'ordre 3 qui vont être multipliés à des polynômes d'ordre 2.

Dans notre expression de départ (C.36), les premières et secondes intégrales (C.37) (C.38) ont été résolues. La première est égale à l'entropie d'une loi normale bivariée, la seconde est nulle. Il nous reste donc à résoudre la troisième intégrale (C.39) définie par :

$$\begin{aligned} & - \oint_{\mathbb{R}^2} \phi_x(\mathbf{u}) \left(B_x(\mathbf{u}) + \frac{(B_x(\mathbf{u}))^2}{2} \right) d\mathbf{u} \\ & = - \oint_{\mathbb{R}^2} \phi_x(\mathbf{u}) B_x(\mathbf{u}) d\mathbf{u} \end{aligned} \tag{C.50}$$

$$- \frac{1}{2} \oint_{\mathbb{R}^2} \phi_x(\mathbf{u}) (B_x(\mathbf{u}))^2 d\mathbf{u} \tag{C.51}$$

Toujours par propriété d'orthogonalité des polynômes d'Hermite, la première intégrale (C.50) est nulle (on rajoute $H_{00}(\mathbf{u}) = 1$ comme précédemment en facteur dans l'intégrale).

Il nous reste donc à développer la seconde intégrale (C.51) :

$$\begin{aligned} & - \frac{1}{2} \oint_{\mathbb{R}^2} \phi_x(\mathbf{u}) (B_x(\mathbf{u}))^2 d\mathbf{u} \\ & = - \frac{1}{2} \oint_{\mathbb{R}^2} \phi_x(\mathbf{u}) \left(\frac{1}{3!} (\kappa^{1,1,1} H_{111}(\mathbf{u}) + 3\kappa^{1,1,2} H_{112}(\mathbf{u}) + 3\kappa^{1,2,2} H_{122}(\mathbf{u}) + \kappa^{2,2,2} H_{222}(\mathbf{u})) \right)^2 d\mathbf{u} \\ & = - \frac{1}{72} \oint_{\mathbb{R}^2} \phi_x(\mathbf{u}) \left(\kappa^{1,1,1^2} H_{111}(\mathbf{u})^2 + 9\kappa^{1,1,2^2} H_{112}(\mathbf{u})^2 + 9\kappa^{1,2,2^2} H_{122}(\mathbf{u})^2 \right. \\ & \quad + \kappa^{2,2,2^2} H_{222}(\mathbf{u})^2 + 6\kappa^{1,1,1} \kappa^{1,1,2} H_{111}(\mathbf{u}) H_{112}(\mathbf{u}) + 6\kappa^{1,1,1} \kappa^{1,2,2} H_{111}(\mathbf{u}) H_{122}(\mathbf{u}) \\ & \quad + 2\kappa^{1,1,1} \kappa^{2,2,2} H_{111}(\mathbf{u}) H_{222}(\mathbf{u}) + 18\kappa^{1,1,2} \kappa^{1,2,2} H_{112}(\mathbf{u}) H_{122}(\mathbf{u}) \\ & \quad \left. + 6\kappa^{1,1,2} \kappa^{2,2,2} H_{112}(\mathbf{u}) H_{222}(\mathbf{u}) + 6\kappa^{1,2,2} \kappa^{2,2,2} H_{122}(\mathbf{u}) H_{222}(\mathbf{u}) \right) d\mathbf{u} \end{aligned} \tag{C.52}$$

Nous devons résoudre les 10 intégrales de (C.52). Nous allons expliquer la résolution sur une de celles-ci, mais elles se traitent toutes de la même manière. A chaque fois, nous utiliserons la propriété d'orthogonalité des tenseurs d'Hermite (C.41) avec $l = k$. Prenons par exemple :

$$\begin{aligned}
 & - \frac{1}{72} \oint_{\mathbb{R}^2} \phi_{\mathbf{x}}(\mathbf{u}) 18 \kappa^{1,1,2} \kappa^{1,2,2} H_{112}(\mathbf{u}) H_{122}(\mathbf{u}) d\mathbf{u} \\
 & = - \frac{18}{72} \kappa^{1,1,2} \kappa^{1,2,2} \oint_{\mathbb{R}^2} \phi_{\mathbf{x}}(\mathbf{u}) H_{112}(\mathbf{u}) H_{122}(\mathbf{u}) d\mathbf{u}
 \end{aligned} \tag{C.53}$$

D'après (C.41), $\oint_{\mathbb{R}^2} \phi_{\mathbf{x}}(\mathbf{u}) H_{112}(\mathbf{u}) H_{122}(\mathbf{u}) d\mathbf{u} = \kappa_{1,1} \kappa_{1,2} \kappa_{2,2} [3!]$, avec κ_{i_1, i_2} un élément de la matrice inverse de covariance $\boldsymbol{\kappa}_{\mathbf{x}}^{-1} = \frac{1}{1-\rho^2} \begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix}$, et $[3!]$ indiquant les 6 permutations possibles des indices données par :

$$(11, 12, 22); (11, 12, 22); (12, 11, 22); (12, 12, 21); (12, 11, 22); (12, 12, 21)$$

On a donc 4 permutations de type (11,12,22) et 2 de type (12,12,21). D'après la matrice $\boldsymbol{\kappa}_{\mathbf{x}}^{-1}$, on peut donc écrire :

$$\begin{aligned}
 & - \frac{18}{72} \kappa^{1,1,2} \kappa^{1,2,2} \oint_{\mathbb{R}^2} \phi_{\mathbf{x}}(\mathbf{u}) H_{112}(\mathbf{u}) H_{122}(\mathbf{u}) d\mathbf{u} \\
 & = - \frac{18}{72} \kappa^{1,1,2} \kappa^{1,2,2} \kappa_{1,1} \kappa_{1,2} \kappa_{2,2} [3!] \\
 & = - \frac{1}{4} \kappa^{1,1,2} \kappa^{1,2,2} (4 \kappa_{1,1} \kappa_{1,2} \kappa_{2,2} + 2 \kappa_{1,2} \kappa_{1,2} \kappa_{2,1}) \\
 & = \frac{1}{4} \kappa^{1,1,2} \kappa^{1,2,2} \left[\frac{4\rho}{(1-\rho^2)^3} + \frac{2\rho^3}{(1-\rho^2)^3} \right] \\
 & = \frac{4\rho + 2\rho^3}{4(1-\rho^2)^3} \kappa^{1,1,2} \kappa^{1,2,2}
 \end{aligned} \tag{C.54}$$

En annexe D, nous donnons les résultats de tous les calculs issus de la propriété (C.41) qui ont été nécessaires au cours de ce travail. Il découle de ces calculs le résultat des 10 intégrales de (C.52) qui s'écrivent :

$$\begin{aligned}
& - \frac{1}{72} \oint_{\mathbb{R}^2} \phi_{\mathbf{x}}(\mathbf{u}) \kappa^{1,1,1^2} H_{111}(\mathbf{u})^2 d\mathbf{u} = -\frac{1}{12(1-\rho^2)^3} \kappa^{1,1,1^2} \\
& - \frac{1}{72} \oint_{\mathbb{R}^2} \phi_{\mathbf{x}}(\mathbf{u}) 9\kappa^{1,1,2^2} H_{112}(\mathbf{u})^2 d\mathbf{u} = -\frac{1+2\rho^2}{4(1-\rho^2)^3} \kappa^{1,1,2^2} \\
& - \frac{1}{72} \oint_{\mathbb{R}^2} \phi_{\mathbf{x}}(\mathbf{u}) 9\kappa^{1,2,2^2} H_{122}(\mathbf{u})^2 d\mathbf{u} = -\frac{1+2\rho^2}{4(1-\rho^2)^3} \kappa^{1,2,2^2} \\
& - \frac{1}{72} \oint_{\mathbb{R}^2} \phi_{\mathbf{x}}(\mathbf{u}) \kappa^{2,2,2^2} H_{222}(\mathbf{u})^2 d\mathbf{u} = -\frac{1}{12(1-\rho^2)^3} \kappa^{2,2,2^2} \\
& - \frac{1}{72} \oint_{\mathbb{R}^2} \phi_{\mathbf{x}}(\mathbf{u}) 6\kappa^{1,1,1} \kappa^{1,1,2} H_{111}(\mathbf{u}) H_{112}(\mathbf{u}) d\mathbf{u} = \frac{\rho}{2(1-\rho^2)^3} \kappa^{1,1,1} \kappa^{1,1,2} \\
& - \frac{1}{72} \oint_{\mathbb{R}^2} \phi_{\mathbf{x}}(\mathbf{u}) 6\kappa^{1,1,1} \kappa^{1,2,2} H_{111}(\mathbf{u}) H_{122}(\mathbf{u}) d\mathbf{u} = -\frac{\rho^2}{2(1-\rho^2)^3} \kappa^{1,1,1} \kappa^{1,2,2} \\
& - \frac{1}{72} \oint_{\mathbb{R}^2} \phi_{\mathbf{x}}(\mathbf{u}) 2\kappa^{1,1,1} \kappa^{2,2,2} H_{111}(\mathbf{u}) H_{222}(\mathbf{u}) d\mathbf{u} = \frac{1\rho^3}{6(1-\rho^2)^3} \kappa^{1,1,1} \kappa^{2,2,2} \\
& - \frac{1}{72} \oint_{\mathbb{R}^2} \phi_{\mathbf{x}}(\mathbf{u}) 6\kappa^{1,1,2} \kappa^{2,2,2} H_{112}(\mathbf{u}) H_{222}(\mathbf{u}) d\mathbf{u} = -\frac{\rho^2}{2(1-\rho^2)^3} \kappa^{1,1,2} \kappa^{2,2,2} \\
& - \frac{1}{72} \oint_{\mathbb{R}^2} \phi_{\mathbf{x}}(\mathbf{u}) 6\kappa^{1,2,2} \kappa^{2,2,2} H_{122}(\mathbf{u}) H_{222}(\mathbf{u}) d\mathbf{u} = \frac{\rho}{2(1-\rho^2)^3} \kappa^{1,2,2} \kappa^{2,2,2}
\end{aligned} \tag{C.55}$$

Concernant (C.51) et après quelques réordonnancements des résultats de (C.54) et (C.55), on obtient :

$$\begin{aligned}
& - \frac{1}{2} \oint_{\mathbb{R}^2} \phi_{\mathbf{x}}(\mathbf{u}) (B_{\mathbf{x}}(\mathbf{u}))^2 d\mathbf{u} \\
& = -\frac{1}{12(1-\rho^2)^3} \left[\kappa^{1,1,1^2} + \kappa^{2,2,2^2} + 3\kappa^{1,1,2^2} + 3\kappa^{1,2,2^2} \right. \\
& \quad - 6\rho(\kappa^{1,1,1} \kappa^{1,1,2} + \kappa^{1,2,2} \kappa^{2,2,2} + 2\kappa^{1,1,2} \kappa^{1,2,2}) \\
& \quad + 6\rho^2(\kappa^{1,1,2^2} + \kappa^{1,2,2^2} + \kappa^{1,1,1} \kappa^{1,2,2} + \kappa^{1,1,2} \kappa^{2,2,2}) \\
& \quad \left. - 2\rho^3(\kappa^{1,1,1} \kappa^{2,2,2} + 3\kappa^{1,1,2} \kappa^{1,2,2}) \right]
\end{aligned}$$

Nous avons désormais résolu toutes les intégrales. Le résultat ci-dessus est le résultat de la deuxième intégrale de (C.39), la première étant nulle. Cette intégrale (C.39) est la troisième intégrale de (C.36). Dans (C.36), la première intégrale est égale à l'entropie d'une distribution normale bivariée, la deuxième est nulle, et la troisième est celle que nous venons de résoudre. Nous pouvons donc finalement écrire :

$$\begin{aligned}
& H_{3E}(\mathbf{x}) = 1 + \log(2\pi) + 0.5 \log(1 - \rho^2) \\
& - \frac{1}{12(1-\rho^2)^3} \left(\kappa^{1,1,1^2} + \kappa^{2,2,2^2} + 3\kappa^{1,1,2^2} + 3\kappa^{1,2,2^2} \right. \\
& \quad - 6\rho(\kappa^{1,1,1} \kappa^{1,1,2} + \kappa^{1,2,2} \kappa^{2,2,2} + 2\kappa^{1,1,2} \kappa^{1,2,2}) \\
& \quad + 6\rho^2(\kappa^{1,1,2^2} + \kappa^{1,2,2^2} + \kappa^{1,1,1} \kappa^{1,2,2} + \kappa^{1,1,2} \kappa^{2,2,2}) \\
& \quad \left. - 2\rho^3(\kappa^{1,1,1} \kappa^{2,2,2} + 3\kappa^{1,1,2} \kappa^{1,2,2}) \right)
\end{aligned} \tag{C.56}$$

l'approximation de l'entropie conjointe par un développement d'Edgeworth à l'ordre 3 pour des variables centrées-réduites.

C.2.2 Extension de l'approximation à l'ordre 4

Nous avons poussé le développement à l'ordre supérieur. La résolution est très similaire. Cette fois, nous utilisons la définition de l'approximation de la densité de probabilité conjointe par un développement d'Edgeworth à l'ordre 4 donnée par :

$$\begin{aligned} p_{\mathbf{x}}(\mathbf{u}) &= \phi_{\mathbf{x}}(\mathbf{u}) \left(1 + \frac{1}{3!} \kappa^{i_1, i_2, i_3} H_{i_1, i_2, i_3}(\mathbf{u}) + \frac{1}{4!} \kappa^{i_1, i_2, i_3, i_4} H_{i_1, i_2, i_3, i_4}(\mathbf{u}) \right. \\ &\quad \left. + \frac{10}{6!} \kappa^{i_1, i_2, i_3} \kappa^{i_4, i_5, i_6} H_{i_1, i_2, i_3, i_4, i_5, i_6}(\mathbf{u}) \right) + O(N^{-3/2}) \end{aligned} \quad (\text{C.57})$$

où il faut considérer toutes les permutations possibles des indices $i_1, i_2, i_3, i_4, i_5, i_6$ quand ces indices prennent les valeurs 1 ou 2. Par rapport au développement à l'ordre 3, on voit tout de suite que la charge de calcul va être plus importante. Les 8 termes de la somme $\kappa^{i_1, i_2, i_3} H_{i_1, i_2, i_3}(\mathbf{u})$ restent les mêmes qu'à l'ordre 3, mais il faut cette fois considérer les 16 termes de la somme $\kappa^{i_1, i_2, i_3, i_4} H_{i_1, i_2, i_3, i_4}(\mathbf{u})$ et les 64 termes de la somme $\kappa^{i_1, i_2, i_3} \kappa^{i_4, i_5, i_6} H_{i_1, i_2, i_3, i_4, i_5, i_6}(\mathbf{u})$. Le $B_{\mathbf{x}}(\mathbf{u})$ de la partie précédente est alors remplacé dans le calcul par :

$$\begin{aligned} C_{\mathbf{x}}(\mathbf{u}) &= \frac{1}{3!} \left(\kappa^{1,1,1} H_{111}(\mathbf{u}) + 3\kappa^{1,1,2} H_{112}(\mathbf{u}) + 3\kappa^{1,2,2} H_{122}(\mathbf{u}) + \kappa^{2,2,2} H_{222}(\mathbf{u}) \right) \\ &\quad + \frac{1}{4!} \left(\kappa^{1,1,1,1} H_{1111}(\mathbf{u}) + 4\kappa^{1,1,1,2} H_{1112}(\mathbf{u}) + 6\kappa^{1,1,2,2} H_{1122}(\mathbf{u}) \right. \\ &\quad \left. + 4\kappa^{1,2,2,2} H_{1222}(\mathbf{u}) + \kappa^{2,2,2,2} H_{2222}(\mathbf{u}) \right) \\ &\quad + \frac{10}{6!} \left(\kappa^{1,1,1} \kappa^{1,1,1} H_{111111}(\mathbf{u}) + 6\kappa^{1,1,1} \kappa^{1,1,2} H_{111112}(\mathbf{u}) + 15\kappa^{1,1,1} \kappa^{1,2,2} H_{111122}(\mathbf{u}) \right. \\ &\quad + 20\kappa^{1,1,1} \kappa^{2,2,2} H_{111222}(\mathbf{u}) + 15\kappa^{1,1,2} \kappa^{2,2,2} H_{112222}(\mathbf{u}) + 6\kappa^{1,2,2} \kappa^{2,2,2} H_{122222}(\mathbf{u}) \\ &\quad \left. + \kappa^{2,2,2} \kappa^{2,2,2} H_{222222}(\mathbf{u}) \right) \end{aligned} \quad (\text{C.58})$$

La démarche reste la même que précédemment : nous commençons par construire l'approximation de l'entropie conjointe à partir de (C.57). Celle-ci est donnée par :

$$H(\mathbf{x}) \approx - \oint_{\mathbb{R}^2} \phi_{\mathbf{x}}(\mathbf{u}) (1 + C_{\mathbf{x}}(\mathbf{u})) \log [\phi_{\mathbf{x}}(\mathbf{u}) (1 + C_{\mathbf{x}}(\mathbf{u}))] d\mathbf{u} \quad (\text{C.59})$$

Dans cette équation, comme précédemment, le seul terme non nul va être :

$$- \frac{1}{2} \oint_{\mathbb{R}^2} \phi_{\mathbf{x}}(\mathbf{u}) (C_{\mathbf{x}}(\mathbf{u}))^2 d\mathbf{u} \quad (\text{C.60})$$

$$= -\frac{1}{2} \oint_{\mathbb{R}^2} \phi_{\mathbf{x}}(\mathbf{u}) \left[\frac{1}{3!} \left(\kappa^{1,1,1} H_{111}(\mathbf{u}) \right. \right. \quad (\text{C.61})$$

$$+ 3\kappa^{1,1,2} H_{112}(\mathbf{u}) + 3\kappa^{1,2,2} H_{122}(\mathbf{u}) + \kappa^{2,2,2} H_{222}(\mathbf{u}) \Big)$$

$$+ \frac{1}{4!} \left(\kappa^{1,1,1,1} H_{1111}(\mathbf{u}) + 4\kappa^{1,1,1,2} H_{1112}(\mathbf{u}) + 6\kappa^{1,1,2,2} H_{1122}(\mathbf{u}) \right.$$

$$+ 4\kappa^{1,2,2,2} H_{1222}(\mathbf{u}) + \kappa^{2,2,2,2} H_{2222}(\mathbf{u}) \Big)$$

$$+ \frac{10}{6!} \left(\kappa^{1,1,1} \kappa^{1,1,1} H_{111111}(\mathbf{u}) + 6\kappa^{1,1,1} \kappa^{1,1,2} H_{111112}(\mathbf{u}) + 15\kappa^{1,1,1} \kappa^{1,2,2} H_{111122}(\mathbf{u}) \right.$$

$$+ 20\kappa^{1,1,1} \kappa^{2,2,2} H_{111222}(\mathbf{u}) + 15\kappa^{1,1,2} \kappa^{2,2,2} H_{112222}(\mathbf{u}) + 6\kappa^{1,2,2} \kappa^{2,2,2} H_{122222}(\mathbf{u})$$

$$\left. + \kappa^{2,2,2} \kappa^{2,2,2} H_{222222}(\mathbf{u}) \right] d\mathbf{u}$$

dont nous devons calculer explicitement les termes. La méthode est la même que précédemment. Parmi tous ces termes ; nous pouvons distinguer.

- Les carrés et doubles produits à l'ordre 3, que nous déjà avons calculés.
- Les produits croisés entre les termes d'ordre 3 et d'ordre 4, qui vont être nuls par propriété des polynômes d'Hermite. En effet, nous aurons des intégrales de type $\int_{\mathbb{R}^2} \phi_{\mathbf{x}}(\mathbf{u}) \kappa^{i_1, i_2, i_3} H_{i_1 i_2 i_3}(\mathbf{u}) \kappa^{i_1, i_2, i_3, i_4} H_{i_1 i_2 i_3 i_4}(\mathbf{u}) d\mathbf{u}$ qui sont égales à 0 d'après (C.41).
- Les carrés et doubles produits à l'ordre 4, qu'il nous reste à calculer.

Nous ne présenterons pas ici tous les détails calculatoires, la démarche étant la même que pour le calcul à l'ordre 3. Cependant, les calculs issus de la propriété des tenseurs d'Hermite qui sont nécessaires pour la résolution de ces intégrales sont donnés en annexe D. Finalement, nous obtenons une approximation de l'entropie conjointe à partir d'un développement d'Edgeworth à l'ordre 4 pour des variables centrées-réduites corrélées donnée par :

$$\begin{aligned}
H_{4E}(\mathbf{x}) = H_{3E}(\mathbf{x}) &- \frac{1}{48(1-\rho^2)^4} \left(\kappa^{1,1,1,1^2} + 4\kappa^{1,1,1,2^2} + 6\kappa^{1,1,2,2^2} + 4\kappa^{1,2,2,2^2} + \kappa^{2,2,2,2^2} \right. \\
&- 8\rho \left(\kappa^{1,1,1,1} \kappa^{1,1,1,2} + 3\kappa^{1,1,1,2} \kappa^{1,1,2,2} + 3\kappa^{1,1,2,2} \kappa^{1,2,2,2} + \kappa^{1,2,2,2} \kappa^{2,2,2,2} \right) + 12\rho^2 \left(\kappa^{1,1,1,2^2} \right. \\
&+ 2\kappa^{1,1,2,2^2} + \kappa^{1,2,2,2^2} + \kappa^{1,1,1,1} \kappa^{1,1,2,2} + 2\kappa^{1,1,1,2} \kappa^{1,2,2,2} + \kappa^{1,1,2,2} \kappa^{2,2,2,2} \Big) - 8\rho^3 \left(\kappa^{1,1,1,1} \kappa^{1,2,2,2} \right. \\
&+ 3\kappa^{1,1,1,2} \kappa^{1,1,2,2} + \kappa^{1,1,1,2} \kappa^{2,2,2,2} + 3\kappa^{1,1,2,2} \kappa^{1,2,2,2} \Big) + 2\rho^4 \left(3\kappa^{1,1,2,2^2} + \kappa^{1,1,1,1} \kappa^{2,2,2,2} \right. \\
&\left. \left. + 4\kappa^{1,1,1,2} \kappa^{1,2,2,2} \right) \right) \\
&- \frac{10}{144(1-\rho^2)^6} \left(\kappa^{1,1,1,1^4} + 6\kappa^{1,1,1,2} \kappa^{1,1,2,2} + 15\kappa^{1,1,1,2} \kappa^{1,2,2,2} + 20\kappa^{1,1,1,2} \kappa^{2,2,2,2} + 15\kappa^{1,1,2,2} \kappa^{2,2,2,2} \right. \\
&+ 6\kappa^{1,2,2,2} \kappa^{2,2,2,2} + \kappa^{2,2,2,4} - 12\rho \left(\kappa^{1,1,1,1^3} \kappa^{1,1,2} + 10\kappa^{1,1,1,2} \kappa^{1,2,2} \kappa^{2,2,2} + 5\kappa^{1,1,1,2} \kappa^{1,1,2} \kappa^{1,2,2} \right. \\
&+ 10\kappa^{1,1,1} \kappa^{1,1,2} \kappa^{2,2,2,2} + 5\kappa^{1,1,2} \kappa^{1,2,2} \kappa^{2,2,2,2} + \kappa^{1,2,2} \kappa^{2,2,2,3} \Big) + 30\rho^2 \left(\kappa^{1,1,1,2} \kappa^{1,1,2,2} + 4\kappa^{1,1,1,2} \kappa^{1,2,2,2} \right. \\
&+ 6\kappa^{1,1,1,2} \kappa^{2,2,2,2} + 4\kappa^{1,1,2,2} \kappa^{2,2,2,2} + \kappa^{1,2,2,2} \kappa^{2,2,2,2} + \kappa^{1,1,1,3} \kappa^{1,2,2} + 4\kappa^{1,1,1,2} \kappa^{1,1,2} \kappa^{2,2,2} \\
&+ 6\kappa^{1,1,1} \kappa^{1,1,2} \kappa^{1,2,2} \kappa^{2,2,2} + 4\kappa^{1,1,1} \kappa^{1,2,2} \kappa^{2,2,2,2} + \kappa^{1,1,2} \kappa^{2,2,2,3} \Big) - 40\rho^3 \left(\kappa^{1,1,1,3} \kappa^{2,2,2} \right. \\
&+ 9\kappa^{1,1,1,2} \kappa^{1,2,2} \kappa^{2,2,2} + 3\kappa^{1,1,1,2} \kappa^{1,1,2} \kappa^{1,2,2} + 3\kappa^{1,1,1} \kappa^{1,1,2,2} \kappa^{2,2,2} + 9\kappa^{1,1,1} \kappa^{1,1,2} \kappa^{2,2,2,2} \\
&+ 3\kappa^{1,1,1} \kappa^{1,2,2,2} \kappa^{2,2,2} + 3\kappa^{1,1,2} \kappa^{1,2,2} \kappa^{2,2,2,2} + \kappa^{1,1,1} \kappa^{2,2,2,3} \Big) + 30\rho^4 \left(3\kappa^{1,1,1,2} \kappa^{1,2,2,2} + 6\kappa^{1,1,1,2} \kappa^{2,2,2,2} \right. \\
&+ 3\kappa^{1,1,2,2} \kappa^{2,2,2,2} + 5\kappa^{1,1,1,2} \kappa^{1,1,2} \kappa^{2,2,2} + 10\kappa^{1,1,1} \kappa^{1,1,2} \kappa^{1,2,2} \kappa^{2,2,2} + 5\kappa^{1,1,1} \kappa^{1,2,2} \kappa^{2,2,2,2} \Big) \\
&- 12\rho^5 \left(5\kappa^{1,1,1} \kappa^{1,1,2,2} \kappa^{2,2,2} + 11\kappa^{1,1,1} \kappa^{1,1,2} \kappa^{2,2,2,2} + 11\kappa^{1,1,1,2} \kappa^{1,2,2} \kappa^{2,2,2} + 5\kappa^{1,1,1} \kappa^{1,2,2,2} \kappa^{2,2,2} \right) \\
&\left. + 2\rho^6 \left(11\kappa^{1,1,1,2} \kappa^{2,2,2,2} + 21\kappa^{1,1,1} \kappa^{1,1,2} \kappa^{1,2,2} \kappa^{2,2,2} \right) \right) \tag{C.62}
\end{aligned}$$

Bibliographie

- [Collignon et al., 1995] Collignon, A., Vandermeulen, D., Suetens, P., and Marchal, G. (1995). 3D Multi-Modality Medical Image Registration Using Feature Space Clustering. *Lecture Notes in Computer Science*, 905 :195–204.
- [Comon, 1994] Comon, P. (1994). Independent component analysis, a new concept? *Signal Processing*, 36(3) :287–314.
- [Studholme and Hawkes, 1995] Studholme, C. Hill, D. L. G. and Hawkes, D. J. (1995). Multiresolution voxel similarity measures for MR-PET registration. In Bizais, Y. and Barillot, C., editors, *Information processing in medical imaging*, Computational imaging and vision, pages 287–298.
- [Van Hulle, 2005] Van Hulle, M. (2005). Multivariate edgeworth-based entropy estimation. *IEEE Workshop on Machine Learning for Signal Processing*, pages 311–316.

Annexe D

Formulaire : propriétés d'orthogonalité des polynômes et tenseurs d'Hermite

Nous donnons ici les résultats issus des propriétés d'orthogonalité des polynômes et tenseurs d'Hermite, qui peuvent être montrées par intégration par parties. Certaines d'entre-elles peuvent être trouvées dans la littérature ([Comon, 1994], p.307, équations (A.13), (A.14), (A.15) et [Amari et al., 1996], p.759, équations (4) et (5)), mais la plupart ont été développées pour la nécessité des développements menés au cours de cette thèse.

Propriétés des polynômes d'Hermite

Ces formules sont issues des identités :

$$\int_{\mathbb{R}} u^{2k+1} \phi(u) du = 0$$

et

$$\int_{\mathbb{R}} u^{2k} \phi(u) du = 1.3...(2k-1)$$

qui découlent de la définition des moments d'une loi normale centrée-réduite. D'une manière générale, nous aurons :

$$\int_{\mathbb{R}} \phi_x(u) (H_{i_1}(u)) \dots (H_{i_n}(u)) du = 0$$

si $\sum_{j=1}^n i_j$ est impair. Dans le cas contraire, cette équation peut être calculée explicitement. Ainsi, nous avons calculé les formules utilisées au cours de la thèse, qui sont données par :

$$\begin{aligned} \int_{\mathbb{R}} \phi_x(u) (H_3(u))^2 du &= 3! \\ \int_{\mathbb{R}} \phi_x(u) (H_4(u))^2 du &= 4! \\ \int_{\mathbb{R}} \phi_x(u) (H_6(u))^2 du &= 6! \end{aligned}$$

$$\begin{aligned}
\int_{\mathbb{R}} \phi_x(u) H_4(u) H_6(u) du &= 0 \\
\int_{\mathbb{R}} \phi_x(u) H_3(u) H_4(u) H_5(u) du &= 12.5! \\
\int_{\mathbb{R}} \phi_x(u) (H_4(u))^3 du &= 12^3 \\
\int_{\mathbb{R}} \phi_x(u) (H_6(u))^3 du &= 100.4!.6! \\
\int_{\mathbb{R}} \phi_x(u) (H_3(u))^2 H_4(u) du &= (3!)^3 \\
\int_{\mathbb{R}} \phi_x(u) (H_3(u))^2 H_6(u) du &= 6! \\
\int_{\mathbb{R}} \phi_x(u) (H_4(u))^2 H_6(u) du &= 16.6! \\
\int_{\mathbb{R}} \phi_x(u) (H_5(u))^2 H_6(u) du &= 200.6! \\
\int_{\mathbb{R}} \phi_x(u) H_4(u) (H_5(u))^2 du &= 20.6! \\
\int_{\mathbb{R}} \phi_x(u) H_4(u) (H_6(u))^2 du &= 5.(3!)^2.6! \\
\int_{\mathbb{R}} \phi_x(u) H_3(u) H_4(u) H_5(u) H_6(u) du &= 211.2^2.(3!)^2.5! \\
\int_{\mathbb{R}} \phi_x(u) (H_3(u))^4 du &= 93.(3!)^2
\end{aligned}$$

Propriétés des tenseurs d'Hermite

De la même manière, nous nous sommes dotés des formules explicites issues de la propriété d'orthogonalité des tenseurs d'Hermite que nous rappelons :

$$\int H_{i_1, \dots, i_k}(\mathbf{u}) H_{j_1, \dots, j_\ell}(\mathbf{u}) \phi_{\mathbf{x}}(\mathbf{u}) d\mathbf{u} = \begin{cases} [k!] \kappa_{i_1, j_1}^{\mathbf{x}} \dots \kappa_{i_k, j_k}^{\mathbf{x}} & \text{si } \ell = k \\ 0 & \text{autrement} \end{cases} \quad (\text{D.1})$$

où $\kappa_{i_1, j_1}^{\mathbf{x}}$ sont les éléments de la matrice inverse de covariance $\boldsymbol{\kappa}_{\mathbf{x}}^{-1} = \frac{1}{1-\rho^2} \begin{pmatrix} 1 & -\rho \\ -\rho & 1 \end{pmatrix}$. En notant, comme au chapitre 3 $\boldsymbol{\kappa}_{\mathbf{x}} = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$ la matrice de covariance de \mathbf{x} et $\rho = \kappa^{1,2}$ la covariance entre $x^{(1)}$ and $x^{(2)}$, on peut écrire pour des VAs centrées réduites :

$$\begin{aligned} \int_{\mathbb{R}^2} \phi_{\mathbf{x}}(\mathbf{u}) (H_{111}(\mathbf{u}))^2 d\mathbf{u} &= \frac{6}{(1-\rho^2)^3} \\ \int_{\mathbb{R}^2} \phi_{\mathbf{x}}(\mathbf{u}) (H_{222}(\mathbf{u}))^2 d\mathbf{u} &= \frac{6}{(1-\rho^2)^3} \\ \int_{\mathbb{R}^2} \phi_{\mathbf{x}}(\mathbf{u}) (H_{112}(\mathbf{u}))^2 d\mathbf{u} &= \frac{2+4\rho^2}{(1-\rho^2)^3} \\ \int_{\mathbb{R}^2} \phi_{\mathbf{x}}(\mathbf{u}) (H_{122}(\mathbf{u}))^2 d\mathbf{u} &= \frac{2+4\rho^2}{(1-\rho^2)^3} \\ \int_{\mathbb{R}^2} \phi_{\mathbf{x}}(\mathbf{u}) H_{111}(\mathbf{u}) H_{112}(\mathbf{u}) d\mathbf{u} &= -\frac{6\rho}{(1-\rho^2)^3} \\ \int_{\mathbb{R}^2} \phi_{\mathbf{x}}(\mathbf{u}) H_{111}(\mathbf{u}) H_{122}(\mathbf{u}) d\mathbf{u} &= \frac{6\rho^2}{(1-\rho^2)^3} \\ \int_{\mathbb{R}^2} \phi_{\mathbf{x}}(\mathbf{u}) H_{111}(\mathbf{u}) H_{222}(\mathbf{u}) d\mathbf{u} &= -\frac{6\rho^3}{(1-\rho^2)^3} \\ \int_{\mathbb{R}^2} \phi_{\mathbf{x}}(\mathbf{u}) H_{112}(\mathbf{u}) H_{122}(\mathbf{u}) d\mathbf{u} &= -\frac{4\rho+2\rho^3}{(1-\rho^2)^3} \\ \int_{\mathbb{R}^2} \phi_{\mathbf{x}}(\mathbf{u}) H_{112}(\mathbf{u}) H_{222}(\mathbf{u}) d\mathbf{u} &= \frac{6\rho^2}{(1-\rho^2)^3} \\ \int_{\mathbb{R}^2} \phi_{\mathbf{x}}(\mathbf{u}) H_{122}(\mathbf{u}) H_{222}(\mathbf{u}) d\mathbf{u} &= -\frac{6\rho}{(1-\rho^2)^3} \\ \int_{\mathbb{R}^2} \phi_{\mathbf{x}}(\mathbf{u}) (H_{1111}(\mathbf{u}))^2 d\mathbf{u} &= \frac{24}{(1-\rho^2)^4} \\ \int_{\mathbb{R}^2} \phi_{\mathbf{x}}(\mathbf{u}) (H_{1112}(\mathbf{u}))^2 d\mathbf{u} &= \frac{(6+18\rho^2)}{(1-\rho^2)^4} \\ \int_{\mathbb{R}^2} \phi_{\mathbf{x}}(\mathbf{u}) (H_{1122}(\mathbf{u}))^2 d\mathbf{u} &= \frac{(4+16\rho^2+4\rho^4)}{(1-\rho^2)^4} \\ \int_{\mathbb{R}^2} \phi_{\mathbf{x}}(\mathbf{u}) (H_{1222}(\mathbf{u}))^2 d\mathbf{u} &= \frac{(6+18\rho^2)}{(1-\rho^2)^4} \\ \int_{\mathbb{R}^2} \phi_{\mathbf{x}}(\mathbf{u}) (H_{2222}(\mathbf{u}))^2 d\mathbf{u} &= \frac{24}{(1-\rho^2)^4} \end{aligned}$$

$$\begin{aligned}
\int_{\mathbb{R}^2} \phi \mathbf{x}(\mathbf{u}) H_{1111}(\mathbf{u}) H_{1112}(\mathbf{u}) d\mathbf{u} &= -\frac{24\rho}{(1-\rho^2)^4} \\
\int_{\mathbb{R}^2} \phi \mathbf{x}(\mathbf{u}) H_{1111}(\mathbf{u}) H_{1122}(\mathbf{u}) d\mathbf{u} &= \frac{24\rho^2}{(1-\rho^2)^4} \\
\int_{\mathbb{R}^2} \phi \mathbf{x}(\mathbf{u}) H_{1111}(\mathbf{u}) H_{1222}(\mathbf{u}) d\mathbf{u} &= -\frac{24\rho^3}{(1-\rho^2)^4} \\
\int_{\mathbb{R}^2} \phi \mathbf{x}(\mathbf{u}) H_{1111}(\mathbf{u}) H_{2222}(\mathbf{u}) d\mathbf{u} &= \frac{24\rho^4}{(1-\rho^2)^4} \\
\int_{\mathbb{R}^2} \phi \mathbf{x}(\mathbf{u}) H_{1112}(\mathbf{u}) H_{1122}(\mathbf{u}) d\mathbf{u} &= -\frac{12\rho + 12\rho^3}{(1-\rho^2)^4} \\
\int_{\mathbb{R}^2} \phi \mathbf{x}(\mathbf{u}) H_{1112}(\mathbf{u}) H_{1222}(\mathbf{u}) d\mathbf{u} &= \frac{18\rho^2 + 6\rho^4}{(1-\rho^2)^4} \\
\int_{\mathbb{R}^2} \phi \mathbf{x}(\mathbf{u}) H_{1112}(\mathbf{u}) H_{2222}(\mathbf{u}) d\mathbf{u} &= -\frac{24\rho^3}{(1-\rho^2)^4} \\
\int_{\mathbb{R}^2} \phi \mathbf{x}(\mathbf{u}) H_{1122}(\mathbf{u}) H_{1222}(\mathbf{u}) d\mathbf{u} &= -\frac{12\rho + 12\rho^3}{(1-\rho^2)^4} \\
\int_{\mathbb{R}^2} \phi \mathbf{x}(\mathbf{u}) H_{1122}(\mathbf{u}) H_{2222}(\mathbf{u}) d\mathbf{u} &= \frac{24\rho^2}{(1-\rho^2)^4} \\
\int_{\mathbb{R}^2} \phi \mathbf{x}(\mathbf{u}) H_{1222}(\mathbf{u}) H_{2222}(\mathbf{u}) d\mathbf{u} &= -\frac{24\rho}{(1-\rho^2)^4} \\
\int_{\mathbb{R}^2} \phi \mathbf{x}(\mathbf{u}) (H_{111111}(\mathbf{u}))^2 d\mathbf{u} &= \frac{720}{(1-\rho^2)^6} \\
\int_{\mathbb{R}^2} \phi \mathbf{x}(\mathbf{u}) (H_{111112}(\mathbf{u}))^2 d\mathbf{u} &= \frac{120 + 600\rho^2}{(1-\rho^2)^6} \\
\int_{\mathbb{R}^2} \phi \mathbf{x}(\mathbf{u}) (H_{111122}(\mathbf{u}))^2 d\mathbf{u} &= \frac{48 + 384\rho^2 + 288\rho^4}{(1-\rho^2)^6} \\
\int_{\mathbb{R}^2} \phi \mathbf{x}(\mathbf{u}) (H_{111222}(\mathbf{u}))^2 d\mathbf{u} &= \frac{36 + 324\rho^2 + 324\rho^4 + 36\rho^6}{(1-\rho^2)^6} \\
\int_{\mathbb{R}^2} \phi \mathbf{x}(\mathbf{u}) (H_{112222}(\mathbf{u}))^2 d\mathbf{u} &= \frac{48 + 384\rho^2 + 288\rho^4}{(1-\rho^2)^6} \\
\int_{\mathbb{R}^2} \phi \mathbf{x}(\mathbf{u}) (H_{122222}(\mathbf{u}))^2 d\mathbf{u} &= \frac{120 + 600\rho^2}{(1-\rho^2)^6} \\
\int_{\mathbb{R}^2} \phi \mathbf{x}(\mathbf{u}) (H_{222222}(\mathbf{u}))^2 d\mathbf{u} &= \frac{720}{(1-\rho^2)^6} \\
\int_{\mathbb{R}^2} \phi \mathbf{x}(\mathbf{u}) H_{111111}(\mathbf{u}) H_{111112}(\mathbf{u}) d\mathbf{u} &= -\frac{720\rho}{(1-\rho^2)^6} \\
\int_{\mathbb{R}^2} \phi \mathbf{x}(\mathbf{u}) H_{111111}(\mathbf{u}) H_{111122}(\mathbf{u}) d\mathbf{u} &= \frac{720\rho^2}{(1-\rho^2)^6} \\
\int_{\mathbb{R}^2} \phi \mathbf{x}(\mathbf{u}) H_{111111}(\mathbf{u}) H_{111222}(\mathbf{u}) d\mathbf{u} &= -\frac{720\rho^3}{(1-\rho^2)^6} \\
\int_{\mathbb{R}^2} \phi \mathbf{x}(\mathbf{u}) H_{111111}(\mathbf{u}) H_{112222}(\mathbf{u}) d\mathbf{u} &= \frac{720\rho^4}{(1-\rho^2)^6} \\
\int_{\mathbb{R}^2} \phi \mathbf{x}(\mathbf{u}) H_{111111}(\mathbf{u}) H_{122222}(\mathbf{u}) d\mathbf{u} &= -\frac{720\rho^5}{(1-\rho^2)^6}
\end{aligned}$$

$$\begin{aligned}
\int_{\mathbb{R}^2} \phi \mathbf{x}(\mathbf{u}) H_{111111}(\mathbf{u}) H_{222222}(\mathbf{u}) d\mathbf{u} &= \frac{720\rho^6}{(1-\rho^2)^6} \\
\int_{\mathbb{R}^2} \phi \mathbf{x}(\mathbf{u}) H_{111112}(\mathbf{u}) H_{111122}(\mathbf{u}) d\mathbf{u} &= -\frac{240\rho + 480\rho^3}{(1-\rho^2)^6} \\
\int_{\mathbb{R}^2} \phi \mathbf{x}(\mathbf{u}) H_{111112}(\mathbf{u}) H_{111122}(\mathbf{u}) d\mathbf{u} &= \frac{360\rho^2 + 360\rho^4}{(1-\rho^2)^6} \\
\int_{\mathbb{R}^2} \phi \mathbf{x}(\mathbf{u}) H_{111112}(\mathbf{u}) H_{112222}(\mathbf{u}) d\mathbf{u} &= -\frac{480\rho^3 + 240\rho^5}{(1-\rho^2)^6} \\
\int_{\mathbb{R}^2} \phi \mathbf{x}(\mathbf{u}) H_{111112}(\mathbf{u}) H_{122222}(\mathbf{u}) d\mathbf{u} &= \frac{600\rho^4 + 120\rho^6}{(1-\rho^2)^6} \\
\int_{\mathbb{R}^2} \phi \mathbf{x}(\mathbf{u}) H_{111112}(\mathbf{u}) H_{222222}(\mathbf{u}) d\mathbf{u} &= -\frac{720\rho^5}{(1-\rho^2)^6} \\
\int_{\mathbb{R}^2} \phi \mathbf{x}(\mathbf{u}) H_{111122}(\mathbf{u}) H_{111222}(\mathbf{u}) d\mathbf{u} &= -\frac{144\rho + 432\rho^3 + 144\rho^5}{(1-\rho^2)^6} \\
\int_{\mathbb{R}^2} \phi \mathbf{x}(\mathbf{u}) H_{111122}(\mathbf{u}) H_{112222}(\mathbf{u}) d\mathbf{u} &= \frac{288\rho^2 + 384\rho^4 + 48\rho^6}{(1-\rho^2)^6} \\
\int_{\mathbb{R}^2} \phi \mathbf{x}(\mathbf{u}) H_{111122}(\mathbf{u}) H_{122222}(\mathbf{u}) d\mathbf{u} &= -\frac{480\rho^3 + 240\rho^5}{(1-\rho^2)^6} \\
\int_{\mathbb{R}^2} \phi \mathbf{x}(\mathbf{u}) H_{111122}(\mathbf{u}) H_{222222}(\mathbf{u}) d\mathbf{u} &= \frac{720\rho^4}{(1-\rho^2)^6} \\
\int_{\mathbb{R}^2} \phi \mathbf{x}(\mathbf{u}) H_{111222}(\mathbf{u}) H_{112222}(\mathbf{u}) d\mathbf{u} &= -\frac{144\rho + 432\rho^3 + 144\rho^5}{(1-\rho^2)^6} \\
\int_{\mathbb{R}^2} \phi \mathbf{x}(\mathbf{u}) H_{111222}(\mathbf{u}) H_{122222}(\mathbf{u}) d\mathbf{u} &= \frac{360\rho^2 + 360\rho^4}{(1-\rho^2)^6} \\
\int_{\mathbb{R}^2} \phi \mathbf{x}(\mathbf{u}) H_{111222}(\mathbf{u}) H_{222222}(\mathbf{u}) d\mathbf{u} &= -\frac{720\rho^3}{(1-\rho^2)^6} \\
\int_{\mathbb{R}^2} \phi \mathbf{x}(\mathbf{u}) H_{112222}(\mathbf{u}) H_{122222}(\mathbf{u}) d\mathbf{u} &= -\frac{240\rho + 480\rho^3}{(1-\rho^2)^6} \\
\int_{\mathbb{R}^2} \phi \mathbf{x}(\mathbf{u}) H_{112222}(\mathbf{u}) H_{222222}(\mathbf{u}) d\mathbf{u} &= \frac{720\rho^2}{(1-\rho^2)^6} \\
\int_{\mathbb{R}^2} \phi \mathbf{x}(\mathbf{u}) H_{122222}(\mathbf{u}) H_{222222}(\mathbf{u}) d\mathbf{u} &= -\frac{720\rho}{(1-\rho^2)^6}
\end{aligned}$$

Bibliographie

- [Amari et al., 1996] Amari, S., Cichocki, A., and Yang, H. H. (1996). A new learning algorithm for blind signal separation. In *Advances in Neural Information Processing Systems*, pages 757–763. MIT Press.
- [Comon, 1994] Comon, P. (1994). Independent component analysis, a new concept? *Signal Processing*, 36(3) :287–314.

Annexe E

Base de données RIRE

Le projet RIRE (Retrospective Image Registration Evaluation) [West et al., 1997], anciennement nommé Retrospective Registration Evaluation Project (RREP) a vu le jour en 1997. Il est à l'initiative de nombreux chercheurs dans le domaine du recalage d'images médicales et avait pour but initial de comparer les méthodes de recalage multimodales en utilisant une référence, ou gold standard. A l'heure actuelle, cette base de données est toujours utilisée intensivement pour évaluer les nouvelles méthodes de recalage multimodal et pour comparer les méthodes existantes. Son avantage majeur est de pouvoir soumettre ses résultats en ligne pour obtenir ses erreurs de recalage.

La base de données utilisée, appelée « Vanderbilt Database », est constituée de volumes d'images provenant de trois modalités différentes : Scanner (CT), TEP et IRM (MR). Ces données ont été acquises au centre hospitalier universitaire Vanderbilt. Avant chaque examen, quatre marqueurs ont été implantés sur chaque patient, ce qui a permis de construire la gold standard, comme nous les verrons dans la suite.

Acquisition des images

Voici les caractéristiques des trois types d'images utilisés dans l'étude :

- les volumes CT ont une résolution de $512 \times 512 \times (28 - 34)$ pixels. La taille des voxels est de 0.65 mm dans les directions x et y , et de 4.0 mm en z . Ils ont été acquis avec un scanner Siemens DR-H.
- les volumes IRM ont une résolution de $256 \times 256 \times (20 - 26)$ pixels. La taille des voxels est comprise entre 1.25 et 1.28 mm dans les directions x et y , et de 4.0 mm en z . Ils ont été acquis avec un scanner Siemens SP 1.5 Tesla. Trois types d'acquisition ont été menés : pondération T1 avec un temps d'écho de 15 ms et un temps de répétition de 650 ou 800 ms, pondération T2 avec un temps d'écho de 90 ms et un temps de répétition de 2550 ou 3000 ms, densité protonique (PD) avec un temps d'écho de 20 ms et un temps de répétition de 2550 ou 3000 ms.
- les volumes TEP ont une résolution de $128 \times 128 \times 15$ pixels. La taille des voxels est de 2.59 mm dans les directions x et y , et de 8.0 mm en z . Ils ont été acquis avec un scanner Siemens/CTI ECAT 933/08-16.

Les images IRM ont de plus été corrigées des artefacts dus au champ magnétique par une technique conçue par [Chang and Fitzpatrick, 1992]. De nouveaux volumes rectifiés sont ainsi disponibles (MR_T1_rectified, MR_T2_rectified, MR_PD_rectified).

Pour chaque patient, 8 volumes d'images (CT, TEP, MR_T1, MR_T2, MR_PD, MR_T1_rectified, MR_T2_rectified, MR_PD_rectified) sont donc disponibles, et 12 recalages

doivent donc être menés (6 recalages CT-IRM et 6 recalages TEP-IRM).

Construction du gold standard

Des marqueurs ont été utilisés pendant les examens, ce qui permet aux auteurs de l'étude de recalcr les images scanner sur les images IRM, et les images TEP sur les images IRM. Pour cela, ils calculent les paramètres de rotation et de translation qui minimisent la distance euclidienne entre les marqueurs sur les différentes images. Les paramètres trouvés servent alors de paramètres de référence (gold-standard). Ensuite, les marqueurs ont été retirés des images. Pour ce faire, les régions des images contenant les marqueurs ont été délimitées manuellement. Puis une reconstruction approximant l'arrière plan a été réalisée dans les régions délimitées. Une illustration de la démarche est donnée figure E.1.

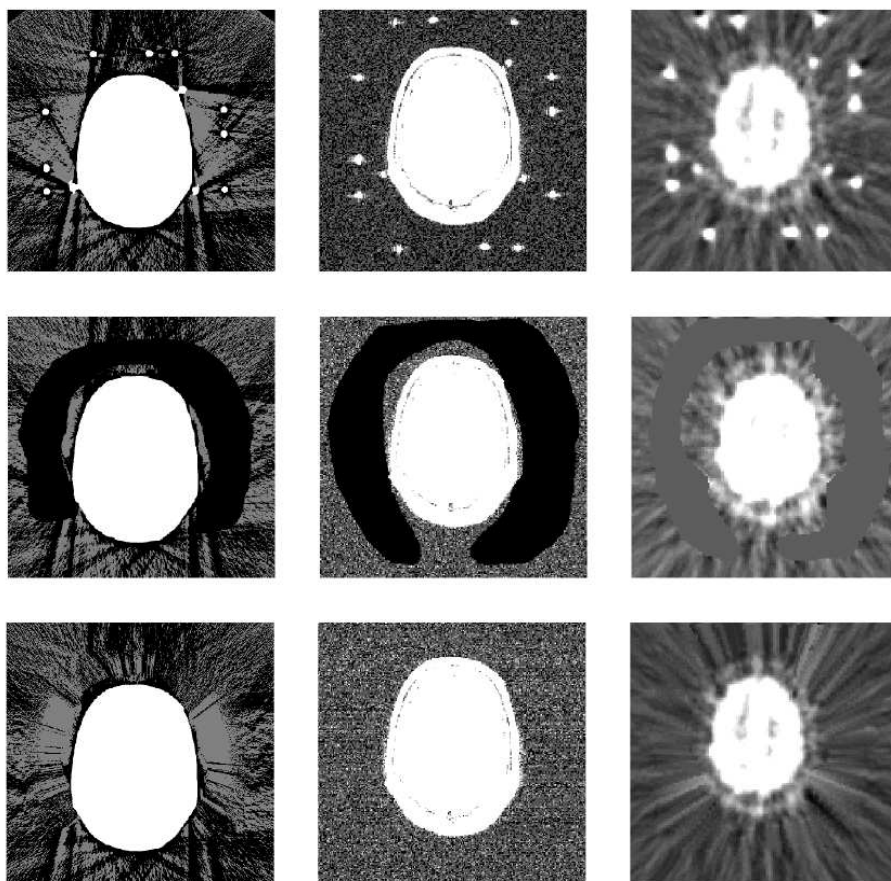


FIGURE E.1 – Suppression des marqueurs utilisés pour le recalage de référence. A gauche, les images scanner. Au milieu les images IRM. A droite, les images TEP. La ligne du haut correspond aux images originales. La ligne du milieu montre les images et les délimitations manuelles des zones contenant ces marqueurs. Enfin, la dernière ligne donne les images reconstruites.

Evaluation des méthodes de recalage

Pour évaluer la qualité d'un recalage (dit recalage test) par rapport au recalage de référence, 10 Volumes d'Interêt (VOI) ont été choisis en collaboration avec des neurologistes et des neurochirurgiens. Ces VOI représentent des zones cliniquement intéressantes. Chaque VOI a été segmenté dans les volumes IRM. Puis une estimation de la qualité du recalage à la position de chaque VOI est calculée de la manière suivante :

- Le pixel central du VOI est trouvé, et sa position est convertie en un point C dans l'image de référence (IRM), en utilisant la taille des voxels, qui est connue, du volume image.
- L'inverse de la transformation de référence (Gold standard) G^{-1} est appliquée à C , donnant la position correspondante $C' = G^{-1}(C)$ dans l'image flottante.
- La transformation du recalage test, R , est appliquée à C' , résultant en $C'' = R(C')$.
- L'erreur du recalage test est ensuite calculée, à la localisation anatomique du VOI, comme la distance Euclidienne entre C et C'' .

Une illustration de cette démarche est donnée figure E.2. Finalement, les erreurs moyenne, médiane et maximale des 10 VOIs est calculée. Ce sont ces 3 valeurs qui permettent l'évaluation quantitative de la méthode de recalage. Dans le chapitre 4, nous nous servons presque exclusivement de l'erreur moyenne, qui nous semble la plus représentative. Mais tous les résultats de la section 4.7 de ce même chapitre sont disponibles en ligne à l'adresse

http://www.insight-journal.org/rire/view_results.php, à mon nom.

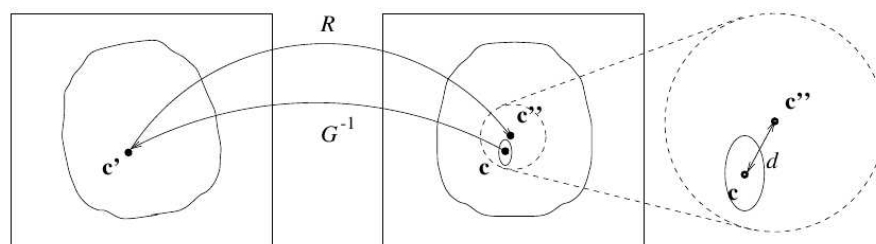


FIGURE E.2 – Calcul de la précision d'un recalage test au niveau d'un VOI (représenté par un cercle pointillé).

Précision du recalage de référence

La validité des évaluations des recalages tests dépend complètement de la précision du recalage de référence. En effet les erreurs de recalage calculées ne sont exactes que si le gold standard fournit un recalage de référence parfait. Et ce n'est évidemment pas le cas. Les auteurs de l'étude ont donc mené des simulations numériques visant à estimer l'erreur commise sur le recalage de référence. Les détails de ces simulations sont données dans [West et al., 1997] et donnent une précision d'approximativement 0.4 mm pour les recalages de référence CT-IRM et de 1.7 mm pour les recalages de référence TEP-IRM. Cette différence s'explique par la taille des voxels des images TEP qui est supérieure à celle des images CT. De plus, si une méthode de recalage testée est aussi efficace que la méthode de référence, elle aura une erreur approximativement égale à $\sqrt{2}$ l'erreur du recalage de référence, soit 0.6 mm pour les recalages de référence CT-IRM et de 2.4 mm pour les recalages de référence TEP-IRM. Ainsi, si une méthode obtient des scores inférieurs à 0.6mm et 2.4mm pour les modalités pré-citées, on peut la considérer comme très efficace.

Bibliographie

- [Chang and Fitzpatrick, 1992] Chang, H. and Fitzpatrick, J. M. (1992). A technique for accurate magnetic resonance imaging in the presence of field inhomogeneities. *IEEE Transactions in Medical Imaging*, 11(3) :319–329.
- [West et al., 1997] West, J., Fitzpatrick, J. M., Wang, M. Y., Dawant, B. M., Maurer, C. R., Kessler, R. M., Maciunas, R. J., Barillot, C., Lemoine, D., Collignon, A., Maes, F., Sumanaweera, T. S., Harkness, B., Hemler, P. F., Hill, D. L. G., Hawkes, D. J., Studholme, C., Maintz, J. B. A., Viergever, M. A., Mal, G., Pennec, X., Noz, M. E., Maguire, G. Q., Pollock, M., Pelizzari, C. A., Robb, R. A., Hanson, D., and Woods, R. P. (1997). Comparison and Evaluation of Retrospective Intermodality Image Registration Techniques. *Journal of Computer Assisted Tomography*, 21(4) :554–566.

Annexe F

Exemple de fichier de paramètres utilisé avec elastiX

Ce fichier a été utilisé pour le recalage non-rigide des volumes scanner des patients atteints d'un cancer de la prostate :

```
// ***** Image Types

(FixedInternalImagePixelType "float")
(FixedImageDimension 3)
(MovingInternalImagePixelType "float")
(MovingImageDimension 3)

// ***** Components

(Registration "MultiResolutionRegistration")
(FixedImagePyramid "FixedRecursiveImagePyramid")
(MovingImagePyramid "MovingRecursiveImagePyramid")
(Interpolator "BSplineInterpolator")
(Metric "Edgeworth3MutualInformationMetric" )
(Optimizer "AdaptiveStochasticGradientDescent")
(ResampleInterpolator "FinalBSplineInterpolator")
(Resampler "DefaultResampler")
(Transform "BSplineTransform")

// ***** Pyramid

// Total number of resolutions
(NumberOfResolutions 3)
(ImagePyramidSchedule 8 8 8 4 4 4 2 2 2 )

// ***** Transform

(FinalGridSpacingInPhysicalUnits 20.0 20.0 20.0)
(GridSpacingSchedule 8.0 4.0 2.0)
(HowToCombineTransforms "Compose")

// ***** Optimizer
```

```
// Maximum number of iterations in each resolution level :
(MaximumNumberOfIterations 200)

(AutomaticParameterEstimation "true")
(UseAdaptiveStepSizes "true")

// ***** Metric
(DerivativeDelta 1)
(FixedImageDimensions 3)

// ***** Several

(WriteTransformParametersEachIteration "false")
(WriteTransformParametersEachResolution "false")
(WriteResultImageAfterEachResolution "false")
(WritePyramidImagesAfterEachResolution "false")
(WriteResultImage "false")
(ShowExactMetricValue "false")
(ErodeMask "false")
(UseDirectionCosines "true")

// ***** ImageSampler

//Number of spatial samples used to compute the mutual information in each resolution level :
(ImageSampler "Random")
(NumberOfSpatialSamples 500)
(NewSamplesEveryIteration "false")
(UseRandomSampleRegion "false")
(SampleRegionSize 50.0 50.0 50.0)
(MaximumNumberOfSamplingAttempts 50)

// ***** Interpolator and Resampler

//Order of B-Spline interpolation used in each resolution level :
(BSplineInterpolationOrder 1)

//Order of B-Spline interpolation used for applying the final deformation :
(FinalBSplineInterpolationOrder 3)

//Default pixel value for pixels that come from outside the picture :
(DefaultPixelValue 0)
```

Table des figures

1.1	Principe du recalage.	18
1.2	Représentations des différents modèles de déformation	28
1.3	Grille de déformation B-Spline	35
1.4	Pondération des champs de déplacement des PDC	35
1.5	Zones d'influence des PDC	36
1.6	Support local des B-Splines	36
1.7	Recherche du pas optimal	39
1.8	Pyramide de transformation	44
1.9	Pyramide gaussienne	46
1.10	Approche bi-pyramidale	46
2.1	Histogramme conjoint	59
2.2	Histogrammes conjoints	59
2.3	Comparaison des estimateurs de densités de probabilité	63
2.4	Manque d'information spatiale apporté par l'IM	69
2.5	Trois histogrammes différents représentant les mêmes données.	73
2.6	Influence du paramètre de lissage des histogrammes à noyau.	73
2.7	Illustration du problème d'estimation pour les dimensions supérieures.	75
2.8	Graphe minimal.	76
3.1	Estimation de l'IM par un développement de Gram-Charlier avec blanchiment.	101
4.1	Critères d'évaluation de la base EMPIRE	108
4.2	Tests en translation	110
4.3	Exemple d'images scanner utilisées	111
4.4	Test de régularité. Rotation de 10 degrés. 50 échantillons	112
4.5	Test de régularité. Rotation de 10 degrés. 100 échantillons	113
4.6	Test de régularité. Rotation de 10 degrés. 500 échantillons	114
4.7	Test de régularité. IMN _H . Rotation de 10 degrés.	115
4.8	Test de régularité. Opposé de 4EIM. Rotation de 10 degrés	117
4.9	Test de régularité. Rotation de 40 degrés. 500 échantillons	118
4.10	Exemples d'images bruitées par un bruit gaussien additif d'écart-type croissant.	119
4.11	Test de régularité. Bruit additif gaussien d'écart-type 1. 50 échantillons	119
4.12	Test de régularité. Bruit additif gaussien d'écart-type 100. 50 échantillons	120
4.13	Test de régularité. Bruit additif gaussien d'écart-type 1. 500 échantillons	120
4.14	Test de régularité. Bruit additif gaussien d'écart-type 100. 500 échantillons	121
4.15	Test de régularité de 3EIM. Bruit additif gaussien. 500 échantillons	122
4.16	Exemples d'images bruitées par un bruit de speckle d'écart-type croissant.	122
4.17	Test de régularité. Bruit de speckle d'écart-type 0,2. 50 échantillons.	123

4.18	Pyramides gaussiennes	126
4.19	Vue d'ensemble des recalages avec l'optimiseur G	128
4.20	Vue d'ensemble des recalages avec l'optimiseur GC	129
4.21	Vue d'ensemble des recalages avec l'optimiseur GS_auto	130
4.22	Vue d'ensemble des recalages avec l'optimiseur GS_manu	131
4.23	Vue d'ensemble des recalages avec l'optimiseur LBFGS	132
4.24	Evaluation du recalage en fonction du nombre d'échantillons pour 3EIM.	133
4.25	Evaluation du recalage en fonction du nombre d'échantillons pour IM _{BS}	134
4.26	Evaluation du recalage en fonction du nombre d'échantillons pour IMN _{BS}	135
4.27	Evaluation du recalage en fonction du nombre d'échantillons pour IM _G	136
4.28	Exemples d'images de la base de données RIRE	136
4.29	Exemple de recalage CT/IRM effectué avec 3EIM.	138
4.30	Exemple de recalage TEP/IRM effectué avec 3EIM.	139
4.31	Difficulté de la segmentation automatique	140
4.32	Approche multi-atlas proposée.	140
4.33	Validation croisée « leave-one-out ».	141
4.34	Coefficient de Dice.	142
4.35	Exemple de masque	143
4.36	Recalage en fonction du nombre d'échantillons	156
4.37	Exemple de résultats de recalage non-rigide aberrant.	157
4.38	Vue d'ensemble des coefficients de Dice obtenus lors de la validation croisée.	157
4.39	Exemples d'images de la base EMPIRE	158
4.40	Evaluation de la qualité des recalages en fonction du nombre d'échantillons.	159
4.41	Evaluation de la qualité des recalages en fonction du nombre d'itérations.	160
4.42	Résultats de la base EMPIRE. Alignement des frontières des Poumons.	161
4.43	Résultats de la base EMPIRE. Alignement des scissures principales.	162
4.44	Résultats de la base EMPIRE. Correspondance entre points annotés.	163
4.45	Résultats de la base EMPIRE. Singularités dans le champ de déformation.	164
B.1	Représentation des B-Splines d'ordre 0 à 3.	174
E.1	Suppression des marqueurs utilisés pour le recalage de référence	198
E.2	Calcul de la précision d'un recalage test au niveau d'un VOI	199

Résumé

Dans le cadre du recalage d'images basé sur l'information d'intensité, l'Information Mutuelle (IM) est couramment utilisée comme mesure de similarité. Cette mesure est en outre particulièrement adaptée au recalage d'images médicales multimodales tri-dimensionnelles. Cependant, les estimateurs de l'IM ont en général une variance élevée et induisent des temps de calcul importants. Au cours de cette thèse, nous nous sommes intéressés aux outils statistiques que sont les cumulants pour construire de nouvelles approximations de l'IM basée sur un développement d'Edgeworth tronqué, le développement d'Edgeworth permettant d'approximer une densité de probabilité à partir de ces cumulants. L'estimée de ces approximations comme mesure de similarité a été évaluée sur données synthétiques et réelles, dans le cadre du recalage rigide et non-rigide d'images médicales multimodales et a été comparée aux estimateurs de référence de l'IM.

Abstract

Mutual Information (MI) is considered as the most common similarity measure in the context of intensity-based image registration. This measure is well-known for its ability to perform tri-dimensional multimodal medical image registration. However, MI's estimators suffer from variance, bias and lead to high computational complexity. During this PhD thesis, we dealt with some statistical tools called cumulants in order to build novel approximations of MI based on Edgeworth expansion. This expansion allows one to approximate a probability density in terms of cumulants. The estimate of these approximations as similarity measure was analyzed in terms of performance on both synthetic and real data, on rigid and non-rigid medical images registration tasks. A comparison with classical estimators of MI was also performed.